The Dissertation Committee for Shu Shen
certifies that this is the approved version of the following dissertation:

Three Essays in Econometrics

Committee:

Stephen Donald, Supervisor

Jason Abrevaya

Richard Chiburis

Thomas Shively

Maxwell Stinchcombe

Three Essays in Econometrics

by

Shu Shen, B.Eco; M.S.Eco

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2011

Dedicated to my parents, Wei Feng and Hong Shen,

and Mr. Hua Li

Acknowledgments

I wish to thank my supervisor Stephen Donald. He has been very supportive and encouraging since the days I began working on this subject. I am extremely grateful for his guidance and help. I am heartily thankful to my committee Jason Abrevaya. His advice and enormous amount of help has been invaluable to my research. I would also like to thank my committee Richard Chiburis, and Maxwell Stinchcombe, for all their comments and help.

I have to thank my parents Wei Feng and Hong Shen for their unconditional support and love. They have supported me throughout all these years I have been away from home working on my degree. I am deeply indebted to my boyfriend Mr. Hua Li who is incredibly understanding and patient. I always have his support whenever I feel frustrated and discouraged. I also thank senior students Yu-Chin Hsu and Tzu-Ying Chen for their amazing help and fellow students Yizao Liu and Piin-Hueih Chiang for their friendship and support.

Finally, I have to thank for the financial support from the Economics department. Without the financial assistance, this research would have been impossible.

Three Essays in Econometrics

Publication No. ____________________

Shu Shen, Ph.D.
The University of Texas at Austin, 2011

Supervisor: Stephen Donald

My dissertation includes three essays that examine or relax classical restrictive assumptions used in econometrics estimation methods. The first chapter proposes methods for examining how a response variable is influenced by a covariate. Rather than focusing on the conditional mean I consider a test of whether a covariate has an effect on the entire conditional distribution of the response variable given the covariate and other conditioning variables. This type of analysis is useful in situations where the econometrician or policy maker is interested in knowing whether a variable or policy would improve the distribution of the response outcomes in a stochastic dominance sense. The response variable is assumed to be continuous, while both discrete and continuous covariate cases are considered. I derive the asymptotic distribution of the test statistics and show that they have simple known asymptotic distributions under the null by using and extending conditional empirical process results given by Horvath and Yandell (1988). Monte Carlo experiments are

conducted, and the tests are shown to have good small sample behavior. The tests are applied to a study on father's labor supply.

The second chapter is based on previous joint work with Jason Abrevaya. It considers estimation of censored panel-data models with individual-specific slope heterogeneity. The slope heterogeneity may be random (random-slopes model) or related to covariates (correlated-random-slopes model). Maximum likelihood and censored least-absolute deviations estimators are proposed for both models. Specification tests are provided to test the slope-heterogeneity models against nested alternatives. The proposed estimators and tests are used for an empirical study of Dutch household portfolio choice. Strong evidence of correlated random slopes for the age variables is found, indicating that the age profile of portfolio adjustment varies significantly with other household characteristics.

The third chapter proposes specification tests in models with endogenous covariates. In empirical studies, econometricians often have little information on the functional form of the structural model, regardless of whether covariates in model are exogenous or endogenous. In this chapter, I propose tests for restricted structural model specifications with endogenous covariates against the fully nonparametric alternative. The restricted model specifications include the nonparametric specification with a restricted set of covariates, the semiparametric single index specification and the parametric linear specification. Test statistics are leave-one-out type kernel U-statistic as used in Fan and Lee (1996). They are constructed using the idea of the control

function approach. Monte Carlo results are provided and tests are shown to have reasonable small sample behavior.

Table of Contents

Acknowledgments **v**

Abstract **vi**

List of Tables **xi**

List of Figures **xii**

Chapter 1. Tests for Distributional Partial Effects **1**

1.1 Introduction 1

1.2 Distributional Hypotheses 6

1.3 The Benchmark Tests: Statistics and Asymptotic Properties . 10

1.3.1 One-sided Nonparametric Testing: the Discrete Covariate Case 11

1.3.2 One-sided Nonparametric Testing: the Continuous Covariate Case 20

1.4 Useful Extensions 25

1.4.1 Single-Index Tests for DPEs 25

1.4.2 Higher Order Stochastic Dominance Tests for DPEs . . 31

1.4.3 Two-sided Tests for DPEs 35

1.5 Monte Carlo Results 37

1.6 Empirical Example 43

1.7 Conclusion 48

Chapter 2. Estimation of Censored Panel-data Models with Slope Heterogeneity **50**
2.1 Introduction 50
2.2 Models and Estimators 52
2.2.1 Maximum Likelihood Estimation 52
Random-Slopes Tobit Models 54
Correlated-random-slopes Tobit Models 57
Partial Effect Estimators 61
Specification Tests 62
2.2.2 Censored Least Absolute Deviation Estimation 64
2.3 Application: Household portfolio decisions 68

Chapter 3. Kernel-based Specification Tests for Models with Endogenous Covariates **82**
3.1 Introduction 82
3.2 Benchmark Test for Nonparametric Restrictions 85
3.2.1 Motivation 85
3.2.2 Test Statistic and Asympotic Property 87
3.3 Parametric and Semiparametric Specification Tests 95
3.3.1 Single Index Models with Endogenous Covariates 95
3.3.2 Parametric Models with Endogenous Covariates 98
3.4 Monte Carlo Experiments 100
3.5 Conclusion 104

Appendices **106**

Appendix A. Chapter 1 Appendix **107**

Appendix B. Chapter 3 Appendix **121**

Bibliography **127**

List of Tables

1.1 Rejection Proportions of DPE Tests: Discrete Covariate Case . 40
1.2 Rejection Proportions of DPE Tests: Discrete Covariate Case . 42

2.1 Descriptive statistics 71
2.2 Portfolio-composition regression results 74
2.3 Standard-deviation estimates for the pooled and random slopes Tobit models 76
2.4 Average partial effect estimates for Tobit models 79
2.5 Mundlak coefficient estimates for the correlated random slopes models 81

3.1 Rejection Proportions of Nonparametric and Semiparametric Tests 102
3.2 Rejection Proportions of the Parametric Test 103

List of Figures

1.1 PE Estimates of Children Gender and Other Family Income . 44
1.2 DPE Tests for Children Gender and Other Family Income . . 45
1.3 Rescaled DPE Functions for Other Family Income 47

Chapter 1

Tests for Distributional Partial Effects

1.1 Introduction

Empirical researchers in economics are often interested in the partial effect (PE) of variables, such as those under control of policy-makers, on economic outcomes. To date the most common approach has been to focus on the partial effects on the mean of the outcome variable using regression techniques. An alternative that has become increasingly popular is the use of quantile regression methods (cf. Koenker 2005) that allow one to examine partial effects at various quantiles typically under assumptions that the conditional quantiles are linear. Recent papers by Koenker and Xiao (2002) and Angrist, Chernozhukov and Fernandez-Val (2006) have considered the quantile regression process and considered methods for inference concerning quantile partial effects over a set of quantiles. In this paper we contribute to the literature on examining effects of variables on distributions by defining the *distributional partial effect* (DPE) which gives the partial effect of a variable, which may be discrete or continuous, at different points in the distribution and examine methods for testing uniform hypotheses concerning distributional partial effect functions using nonparametric and semiparametric estimators. The uniform hypotheses we look at in this paper include both one-sided and two-sided null hypotheses. One-sided tests study whether a DPE function has uniform sign, or a covariate has a uniformly positive or negative effect on the conditional distribution of the response variable. Two-sided tests ask whether the DPE function is zero, or the covariate has insignificant impact on the conditional distribution of the response variable. We also consider extensions of the tests that relate to higher order stochastic dominance relationship between the DPE function and the zero functions. Higher order stochastic dominance tests are especially useful when first order one-sided null hypotheses are too strong to be useful in empirical studies.

All of our DPE tests are built upon kernel-based conditional empirical processes, which could be viewed as nonparametric or semiparametric "dual" counterparts of parametric quantile regression processes discussed in Koenker and Xiao (2002). In contrast to their hypotheses which concern the coefficients on an assumed linear quantile function across a set of quantiles, our test concerns the effect of a covariate on the conditional distribution function directly and considers hypotheses that the effect is uniformly positive, negative or zero at all points in the support of the conditional distribution. Also we allow for the conditional distribution to be nonparametric and also consider semiparametric versions. Under the exogeneity assumption of conditioning variables, the distributional partial effect is the partial effect of a covariate on the conditional distribution of a response given that covariate and other conditioning variables and is a function from the support of the response or outcome variable to the real line. Both our tests and those in Koenker and Xiao (2002) are distribution free in the sense that the limit distributions do not depend on nuisance parameters. While tests in Koenker and Xiao (2002) enjoy parametric convergence rates and ours does not, the methods rely on linearity of the quantiles and may therefore be subject to parametric misspecification error as discussed in Angrist, Chernozhukov and Fernandez-Val (2006).

Our DPE tests are also related to a large literature on stochastic dominance in the sense that we formulate our tests by studying the stochastic dominance relationship of DPEs and the zero function. Stochastic dominance tests without covariates include Anderson (1996), Davidson and Duclos (1997, 2000), Barrett and Donald (2003), Horváth, Kokoszka and Zitikis (2006), among others. Other papers that incorporate covariates in restrictive ways include Linton, Maasoumi and Whang (2005) who consider stochastic dominance relations for residuals from a linear regression and Donald and Hsu (2010) who studies stochastic dominance relations of unconditional response distributions with and without treatment using an ignorability assumption which is conditional on covariates. Lee and Whang (2009) considers stochastic dominance relations between conditional treatment and control distributions using the ignorability assumption. Lee, Linton and Whang (2010) and Delgado and Escanciano (2010) study stochastic monotonicity tests, which could be thought of as stochastic dominance tests with a continuous covariate. Our

DPE tests, with the exogeneity assumption on conditioning covariates relaxed to the unconfoundedness assumption, could be viewed as tests for conditional distributional treatment effects as well but are different from Lee and Whang (2009) in the sense that we study partial effects of a covariate on the conditional distribution of the response evaluated at fixed values of conditional variables while their tests have stronger hypotheses focusing on effects of a covariate for all possible values of conditional variables. Similarly, our tests for DPEs of continuous covariates study weaker hypotheses than Lee, Linton and Whang's (2010) and Delgado and Escanciano's (2010) stochastic monotonicity tests. We will discuss the relationship between these various hypotheses after distributional hypotheses are introduced in the next section.

The test statistics considered in this paper are based on kernel estimation techniques and Kolmogorov-Smirnov type of functionals and are shown to converge asymptotically to simple distributions related to changed time Brownian Bridge processes. For both the one-sided and two-sided benchmark tests based on either the nonparametric or semiparametric single-index estimators, the limiting distributions are nuisance parameter free with critical values that are easily tabulated. We also consider extensions of the tests that relate to higher order stochastic dominance relations in which case the limit distributions are not nuisance parameter free for reasons similar to that found in Barrett and Donald (2003) for their higher order tests. In these cases we propose a simple simulation method for obtaining critical values. Our tests being of Kolmogorov-Smirnov type are based on the maximal difference over the support of the distribution and is straightforward to compute in practice. One could use alternative functionals that would yield consistent tests such as an L^1-norm as in Lee and Whang (2010) or an L^2-norm as in Yatchew (2005). As noted in McCaig and Yatchew (2007), Kolmogorov-Smirnov type statistics are expected to have greater power against alternatives that sees a large but short deviation from the null, while the norm type statistics shall be more powerful in detecting small but persistent deviations from the null hypothesis.

The remainder of the paper is organized as follows. In Section 1.2, we introduce the basic notations and hypotheses of interest and also relate the hypotheses to notions of differential or marginal stochastic dominance. In Section 1.3, we propose test statistics for two benchmark tests based on nonparametric estimators and study their asymptotic properties. In section 1.4 we extend benchmark tests to tests based on semiparametric estimators which are less robust but have faster convergence rates. Also in Section 1.4 we extend the tests to higher order stochastic dominance relations and consider two-sided tests. Section 1.5 conducts Monte Carlo experiments to examine the small sample behavior of proposed tests. Section 1.6 contains an empirical example which examines the relationship between children gender and parental income. Section 1.7 concludes the contribution of the paper and suggests directions of future research. Proofs for theorems and propositions are provided in the appendix.

1.2 Distributional Hypotheses

In this section we state the distributional hypotheses of interest when the econometrician or policy maker wants to know the effect of an exogenous variable on the entire distribution of outcomes, and is not focused on a single characteristic such as the mean or median which may hide interesting and important distributional impacts of a change in the exogenous variable. Our benchmark tests are nonparametric one-sided tests that examine the first order stochastic dominance relationship of DPEs and the zero function, or in other words, whether DPEs have uniform sign. We will also discuss semiparametric tests, higher order stochastic dominance tests and two-sided tests in corresponding extension sections. One thing to note is that although exogoneity type assumptions are required for a "causal" interpretation of the DPEs, such assumptions are not required to justify the asymptotic results. We will discuss this more in below.

Let Y and X be random variables where Y is continuous with compact support $\mathcal{Y} \subseteq \mathbb{R}$ and $X = (X_1, X_2)$ is p-dimensional. First, suppose that X_1 is a dummy variable and that X_2 is continuous. Let $\mathcal{W}$ be the set of X_2 values that have positive densities, $\mathcal{W} \subseteq \mathbb{R}^{p-1}$. Under the exogeneity assumption, the DPE of X_1 is the difference between $F_{Y|X}$ evaluated at different X_1 and fixed X_2 values. Let $F_i(y|x_2)$ denote the shorthand notation for $F_{Y|X_1=i,X_2=x_2}(y|x_2)$, $i = 0, 1$. The first null hypothesis we study in this paper is as follows

$$H^1_{0,x_2} : \text{For fixed } x_2 \in \mathcal{W}, F_1(y|x_2) \leq F_0(y|x_2) \text{ for all } y \in \mathcal{Y}.$$

The null hypothesis states that the conditional distribution of Y at $X_1 = 1$, $X_2 = x_2$ first order stochastically dominates that at $X_1 = 0$, $X_2 = x_2$. In other words, for fixed x_2 the probability that Y is less than y is always lower when $X_1 = 1$ than when $X_1 = 0$. If larger values for Y correspond to "better" outcomes, such as when Y represents income or consumption, then this hypothesis corresponds to the idea that conditional on the fact the $X_2 = x_2$ the distribution of outcomes for $X_1 = 1$ is "better" than that for $X_1 = 0$. Mathematically, we have

$$F_1(.|x_2) \leq F_2(.|x_2) \Leftrightarrow \int_{\mathcal{Y}} u(y)dF_1(y|x_2) \geq \int_{\mathcal{Y}} u(y)dF_0(y|x_2)$$
$$\text{for all increasing } u \text{ functions.}$$

The increasing $u(.)$ function could be an individual utility function and the integrals may be interpreted as social welfare functions defined conditionally on $X_2 = x_2$ for those who have $X_1 = 1$ and those who have $X_1 = 0$. As a special case of the above relationship, $u(y) = y$ and H^1_{0,x_2} implies that $E(Y|X_1 = 1, X_2 = x_2) \geq E(Y|X_1 = 0, X_2 = x_2)$, i.e. that the partial effect on the mean of X_1 evaluated at $X_2 = x_2$ is positive. But the converse statement is obviously not true and indeed it may be that the mean effect can be positive but that the distributional effect may differ in different parts of the distribution of Y in ways that could be undesirable. When X_1 is a treatment variable and the ignorability condition is imposed instead of the exogeneity assumption then these have interpretations as treatment effects.

When there is no conditioning variable X_2, the null hypothesis degenerates to the first order stochastic dominance test without conditioning covariates in Barret and Donald (2003). When the uniform inequality is formulated for all possible X_2 values instead of at given fixed values, the null hypothesis becomes the one in Lee and Whang (2009). Therefore, Lee and Whang (2009) looks at a stronger test than ours: if H^1_{0,x_2} is false, Lee and Whang (2009)'s null is certainly false; but not vice versa. Although Lee and Whang (2009)'s test has some nice features, such as power properties typical of parametric tests, in applications with many covariates the null hypothesis for their test is likely to be too strong and their test computationally too demanding to make it useful in practice. In contrast, our test enables econometricians to study the DPE of X_1 evaluated at any interesting values of conditioning X_2. For example, if a covariate is expected to affect the response variable through two mechanisms in opposite directions, it might be possible that one mechanism dominates for some X_2 values and the other for the rest. Then the econometrician may want to evaluate the DPE at different X_2 values.

Now consider the case where the covariate X_1 is continuous. Under the exogeneity of X, the DPE function of X_1 is the partial derivative of $F_{Y|X}$ with respect to X_1. Let $\mathcal{X}$ be the set of X values that have positive densities. The null hypothesis that we are interested in tests whether the DPE function has uniformly negative sign.

$$H^2_{0,x} : \text{For fixed } x \in \mathcal{X}, \frac{\partial}{\partial x_1} F_{Y|X=x}(y|x) \leq 0 \text{ for all } y \in \mathcal{Y}.$$

If $H^2_{0,x}$ is true, the DPE of X_1 evaluated at $X = x$ stochastically dominates the zero function. In other words, at the value x a marginal increase in X_1, holding

X_2 fixed, decreases the probability that Y is less than y uniformly across the support. For similar reasons to the discrete case when $H^2_{0,x}$ holds and larger values for Y correspond to "better" outcomes then a marginal increase in X_1 has an unambiguously positive effect on outcomes in the distributional sense. Mathematically,

$$\frac{\partial}{\partial x_1}F(.|x) \leq 0 \Leftrightarrow \frac{\partial}{\partial x_1}\int_{\mathcal{Y}} u(y)dF(y|x) = -\int_{\mathcal{Y}} u'(y)\frac{\partial}{\partial x_1}F(y|x)dy \geq 0$$
$$\text{for all increasing } u \text{ functions.}$$

The equality is obtained from integration by parts. If $u(y)$ represents an individual measure of the utility of $Y = y$ then the integral against the conditional distribution of Y given $X = x$ gives a measure of overall welfare conditional on $X = x$. The hypothesis then implies that regardless of the form of $u(y)$ an increase in X_1 yields higher overall welfare conditional on $X = x$. As before this implies an increase in the mean of Y but the converse is not true for reasons discussed above.

The closest existing literatures to our second benchmark test are stochastic monotonicity tests proposed in Lee, Linton and Whang (2010) and Delgado and Escanciano (2010). Their tests are equivalent to one that studies whether a continuous covariate has uniformly positive or negative DPEs over all of its possible values, when the conditioning distribution of the response variable on the single dimensional covariate is differentiable. Neither of their tests could be used for testing DPE of a covariate with multidimensional conditional variables. [1] Therefore, their null hypotheses are much stronger versions of $H^2_{0,x}$ when the dimension of X is equal to 1, or $p = 1$. Also, their tests can be expensive in terms of computational time. In section 1.5 we compare the small sample behavior of our second benchmark test when $p = 1$ with LLW's test and find that our test perform at least as good as theirs in small sample.

1.3 The Benchmark Tests: Statistics and Asymptotic Properties

In this paper, we only consider the case where we have independent data drawn from an identical distribution. The response variable must be continuous while the regressors could be continuous or discrete or a mix of continuous and discrete variables.

Assumption 1.3.1.

1. *$\{Y_i, X_{1i}, X_{2i}\}_{i=1}^n$ are random sample of size n of $\{Y, X_1, X_2\}$, which is characterized by the cumulative distribution function $G(y, x_1, x_2)$.*

2. *Denote the support of Y as $\mathcal{Y} = [0, \bar{y}]$, $\bar{y} < \infty$. Let $X = (X_1, X_2)$.*

We are interested in testing whether the covariate X_1 has uniformly negative DPE at fixed values of other conditioning variables and itself. Both cases of discrete and continuous X_1 variables are studied. The conditioning

[1] Lee, Linton and Whang (2010) also proposed a test for stochastic monotonicity in a vector. The null hypothesis might be viewed as testing whether DPEs of each X element are uniformly negative conditional on all possible X values. Also, it maybe too strong a hypothesis to be useful in empirical studies.

X_2 is not limited to continuous variables, however we will only discuss cases with continuous X_2 since discrete variables can be treated by sample splitting. Though sample splitting is fully nonparametric it may result in a deterioration of the convergence rate of test statistics to be discussed due to the shrinking of the sample size. In section 1.4 we discuss semiparametric extensions of the benchmark tests to get around sample splitting when we have discrete elements in X_2 .

1.3.1 One-sided Nonparametric Testing: the Discrete Covariate Case

First we study the one-sided nonparametric benchmark test for discrete covariates. Without loss of generality, assume further that X_1 is a dummy variable and that the response $Y = H(X, e)$, where H is an unknown structural function and e is the unobservable determinant of Y. The null hypothesis as is formulated in H^1_{0,x_2} focuses on the sign of the difference function $F_{Y|X_1=1,X_2=x_2} - F_{Y|X_1=0,X_2=x_2}$. We first discuss two situations where the hypothesis has a causal interpretation. The first is based on standard exogeneity assumptions, which involves independence assumptions between X and e, while the second is based on assumptions within the potential outcomes framework. For the latter let $Y(i) = H(X_1 = i, X_2, e) \overset{let}{=} H_i(X_2, e)$, be the potential outcomes for $X_1 = i$ where $i = 0, 1$. The following two assumption give situations where the difference function $F_{Y|X_1=1,X_2=x_2} - F_{Y|X_1=0,X_2=x_2}$ has a causal interpretation.

Assumption A. $X \perp e$.

Assumption B. $(Y(0), Y(1)) \perp X_1 \mid X_2$.

Assumption A is the standard exogeneity assumption econometricians use to make causal inference with mean estimation results. It is also sufficient for our distributional estimation analysis to have a causal interpretation as for all $y \in \mathcal{Y}$,

$$
\begin{aligned}
&F_{Y|X_1=1,X_2=x_2}(y|x_2) - F_{Y|X_1=0,X_2=x_2}(y|x_2) \\
&= E_e[1(H_1(x_2, e) \leq y)|X_1 = 1, X_2 = x_2] - E_e[1(H_0(x_2, e) \leq y)|X_1 = 0, X_2 = x_2] \\
&= E_e[1(H_1(x_2, e) \leq y)] - E_e[1(H_0(x_2, e) \leq y)], \\
&= E_e[1(H_1(x_2, e) \leq y) - 1(H_0(x_2, e) \leq y)],
\end{aligned}
$$

where the second equality is from Assumption A. The last expression could be easily understood as the average influence of X_1 (at $X_2 = x_2$) on the event that the response variable is less than some y value over the marginal distribution of the unobservable determinant e. Given that the equality holds for all possible y values, under Assumption A $F_{Y|X_1=1,X_2=x_2}(y) - F_{Y|X_1=0,X_2=x_2}$ gives the unconditional causal effect of X_1 on the distribution of Y when $X_2 = x_2$. We call it the distributional partial effect of X_1.

Econometricians or policy makers sometimes find Assumption A too restrictive to impose in empirical work. Instead, they impose in their model Assumption B, which is the standard unconfoundedness or selection on observables assumption that has been commonly used in the treatment effect literature. Imbens (2004) has a thorough discussion on the meaning of the unconfoundedness assumption in applications. Under Assumption B, $F_{Y|X_1=1,X_2=x_2}(y) - F_{Y|X_1=0,X_2=x_2}$ could be shown equal to an expression indicating the average influence of X_1 on the event that the response variable is less than some y value over the conditional distribution of the unobservable determinant e (conditional on $X_2 = x_2$). In such cases, we call $F_{Y|X_1=1,X_2=x_2}(y) - F_{Y|X_1=0,X_2=x_2}$ conditional distributional treatment effect.

$$
\begin{aligned}
&F_{Y|X_1=1,X_2=x_2}(y|x_2) - F_{Y|X_1=0,X_2=x_2}(y|x_2) \\
&= E_e[1(H_1(x_2,e) \leq y)|X_1 = 1, X_2 = x_2] - E_e[1(H_0(x_2,e) \leq y)|X_1 = 0, X_2 = x_2] \\
&= E_e[1(H_1(x_2,e) \leq y)|X_2 = x_2] - E_e[1(H_0(x_2,e) \leq y)|X_2 = x_2], \\
&= E_e[1(H_1(x_2,e) \leq y) - 1(H_0(x_2,e) \leq y)|X_2 = x_2].
\end{aligned}
$$

The second equality in above follows from the unconfoundedness assumption in B. Notice that the conditional distributional treatment effect averages the effect of X_1 over the conditional distribution of e (conditional on $X_2 = x_2$) while the distributional partial effect averages the effect over the marginal distribution of e. Therefore, although both give causal inference, Assumption A is stronger in that it requires e and X_2 to be independent but estimation and testing results obtained under Assumption B does not have "external validity" in the sense that they shall not be directly used as information for another sample unless econometricians or policy makers are certain that the conditional distribution of the unobserved heterogeneity and controlling variables are the same for the two samples. Also note that in general both partial effects depend on the fixed X_2 value where effects of X_1 are evaluated.

Neither Assumption A nor Assumption B is required for the asymptotic results discussed in the rest of this section. It only provides a situation where the null and alternative have a direct structural interpretation. If it is not valid, the null hypothesis still gives information on the relationship between the covariate and the conditional distribution of the response variable. Examples where the test described above might be interesting could be the effect of children gender (X_1) on the distribution of fathers' labor income (Y) controlling household characteristics (X_2) such as other family income, father's education and so on, or the treatment effect of a training program (X_1) on the distribution of unemployed spell length (Y) controlling the worker's characteristics (X_2) such as age, education, wealth level and so on, or others.

Let $q = p - 1$ be the dimension of the conditioning variable X_2, $q \geq 1$. Assume that the underlying joint distribution of Y and X satisfies the following conditions.

Assumption 1.3.2.

1. $X_1 \in \{0,\ 1\}$. *Let* $\mathcal{W}$ *be the set of* X_2 *values that have positive densities and* $P(X_1 = 1|X_2 = x_2) \in (0, 1)$. $\mathcal{W} \in \mathbb{R}^q$ *is a compact set.*

2. *Let* $G_i(y, x_2) = P[Y \leq y, X_1 = i, X_2 \leq x_2]$ *be the continuous cumulative joint distribution function of* Y *and* X_2 *at* $X_1 = i$*, and let* $s_i(y, x_2)$ *be its corresponding probability distribution function,* $i = 0, 1$.

3. *Let* $g_i(y, x_2) = \int 1(u \leq y)s_i(u, x_2)du$, $i = 0, 1$. $g_i(y, x_2)$ *is uniformly bounded up until second derivatives.*

4. *Let* $f_i(x_2) = g_i(\infty, x_2)$*,* $i = 0, 1$*.* $f_i(x_2)$ *is uniformly bounded up until second derivatives.*

It is obvious that f_0, f_1 are probability density functions of X_2 at $X_1 = 0$ and 1. Let $K : \mathbb{R}^q \to \mathbb{R}$ be the kernel function. Let $n_0 = \sum_{i=1}^n 1(X_{1i} = 0)$, $n_1 = \sum_{i=1}^n 1(X_{1i} = 1)$ the number of observations that have $X_1 = 0$ or 1 respectively. For fixed x_2 values in $\mathcal{W}$, kernel density estimators of $f_0(x_2)$ and $f_1(x_2)$ are defined as follows,

$$\hat{f}_0(x_2) = \frac{1}{n_0 h_0^q} \sum_{i=1}^n 1(X_{1i} = 0) K\left(\frac{X_{2i} - x_2}{h_0}\right),$$
$$\hat{f}_1(x_2) = \frac{1}{n_1 h_1^q} \sum_{i=1}^n 1(X_{1i} = 1) K\left(\frac{X_{2i} - x_2}{h_1}\right),$$

where h_0, h_1 are bandwidths of the two nonparametric estimators. Since $g_i(y, x_2) = P(Y \leq y, X_i = i, X_2 = x_2)$, we know that $g_i(y, x_2)/f_i(x_2) = F_{Y|X_1=i,X_2=x_2}(y|x_2)$, $i = 0, 1$. For simplicity, we use shorthand notation $F_i(y|x_2)$ for the conditional distribution $F_{Y|X_1=i,X_2=x_2}(y|x_2)$ in the rest of the paper. For fixed $y \in \mathcal{Y}$ and $x_2 \in \mathcal{W}$, Nadaraya-Waston kernel estimators of $F_0(y|x_2)$ and $F_1(y|x_2)$ are,

$$\hat{F}_0(y|x_2) = \frac{\sum_{i=1}^n 1(X_{1i} = 0) K\left(\frac{X_{2i} - x_2}{h_0}\right) 1(Y_i \leq y)}{\sum_{i=1}^n 1(X_{1i} = 0) K\left(\frac{X_{2i} - x_2}{h_0}\right)},$$
$$\hat{F}_1(y|x_2) = \frac{\sum_{i=1}^n 1(X_{1i} = 1) K\left(\frac{X_{2i} - x_2}{h_1}\right) 1(Y_i \leq y)}{\sum_{i=1}^n 1(X_{1i} = 1) K\left(\frac{X_{2i} - x_2}{h_1}\right)}.$$

Given the estimators, we propose the following test statistic for testing the null hypothesis H^1_{0,x_2}:

$$\hat{S}_1(x_2) = A_1 \left(\frac{n_1 h_1^q \hat{f}_1(x_2) n_0 h_0^q \hat{f}_0(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)} \right)^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \left(\hat{F}_1(y|x_2) - \hat{F}_0(y|x_2) \right),$$

where $A_1 = (\int K(\phi)^2 d\phi)^{-\frac{1}{2}}$ is a constant that adjusts with the choice of the kernel function. When X_2 is single-dimensional, $A_1 = (5/3)^{\frac{1}{2}}$ for the Epanechnikov kernel function and $(2\sqrt{\pi})^{\frac{1}{2}}$ for the Gaussian kernel function. The limiting distribution of $\hat{S}_1(x_2)$ under the null is characterized based on the asymptotic properties of kernel-based conditional empirical processes developed in the statistics literature. See, among many others, Stute (1986, 1986a) and Horvath and Yandell (1988).

Assumption 1.3.3. *The kernel function K satisfies:*

1. $\int K(\phi) d\phi = 1$, $\int \phi K(\phi) d\phi = 0$;
2. $\sup |K(\phi)| < \infty$, $\int |K(\phi)| d\phi < \infty$, $\int |\phi K(\phi)| d\phi < \infty$;
3. $\mu_2 = \int \phi^2 K(\phi) d\phi < \infty$.

Assumption 1.3.4. *The two bandwidths h_0 and h_1 satisfy the following conditions:*

1. $n_0 h_0^q$, $n_1 h_1^q \to \infty$ *as* $n \to \infty$;
2. $n_0 h_0^{q+4}$, $n_1 h_1^{q+4} \to 0$ *as* $n \to \infty$.

Note that Assumption 1.3.4 implies undersmoothing. The optimal bandwidth in the sense of minimizing Asymptotic Mean Integrated Squared Error (AMISE), which is of order $O\left(n^{-\frac{1}{q+4}}\right)$ (see for example Pagan and Ullah (1999)), does not satisfy the above criteria and will result in some bias in the centering for the limiting distribution of kernel based conditional empirical processes. Bandwidth's that satisfy this assumption should be of order $O\left(n^{-\frac{1}{q+4}+k}\right)$ for some small positive k value. These requirements for the bandwidth are the same as those for asymptotic normality of pointwise Nadaraya-Watson estimators.

Let $D[0,\bar{y}]$ denote the space of all *cadlag* functions ("right continuous with left limits") on $\mathcal{Y} = [0,\bar{y}]$, and $\mathcal{B}(D[0,\bar{y}])$ the generated Borel σ field. For fixed x_2, both kernel based conditional distribution function estimators $\hat{F}_0(.|x_2)$ and $\hat{F}_1(.|x_2)$ are random elements in $(D[0,\bar{y}], \mathcal{B}(D[0,\bar{y}]))$.

Theorem 1.3.1. *Under Assumption 1.3.1-1.3.4, we have that*

$$A_1[n_0 h_0^q \hat{f}_0(x_2)]^{\frac{1}{2}}(\hat{F}_0(.|x_2) - F_0(.|x_2)) \Rightarrow \mathfrak{B}(F_0(.|x_2))$$

in $D([0,\bar{y}])$.

Here $\mathfrak{B}(.)$ is a standard Brownian Bridge Process on the unit interval $[0,1]$. Let $C([0,\bar{y}])$ denote the space of all continuous functions on $[0,\bar{y}]$, $P[\mathfrak{B}\left(F_0(.|x_2)\right) \in C([0,\bar{y}])] = 1$. A corresponding result holds for the process concerning the $X_1 = 1$ observations. We know from Theorem 1.3.1 that for the fixed x_2 value the sequences of conditional distribution estimators $\hat{F}_0(.|x_2)$ and

$\hat{F}_1(.|x_2)$ after proper centering and rescaling converge to scaled changed time Brownian Bridge processes at rate $\sqrt{n_0 h_0^p}$ and $\sqrt{n_1 h_1^p}$ respectively. Notice that $\text{Var}[\mathfrak{B}\left(F_i(y|x_2)\right)] = F_i(y|x_2)[1 - F_i(y|x_2)] = \text{Var}(1(Y \leq y)|X_1 = i, X_2 = x_2)$, for $i = 0, 1$ and all $y \in \mathcal{Y}$. If we fix the y value, Theorem 1.3.1 degenerates to the standard weak convergence result of the pointwise Nadaraya-Waston estimator for $E[1(Y \leq y)|X_1 = i, X_2 = x_2]$, $i = 0, 1$. Asymptotic variances of the estimated processes are influenced by the density of the conditioning variable X_2 at the fixed value x_2, a property inherited from the nature of Nadaraya-Waston estimators. The smaller is the density, the larger is the variance. Therefore, although the test is theoretically applicable to all fixed values of the conditioning variable with positive densities, in empirical studies with limited sample size the test would be more precise if the conditioning variable X_2 is fixed at some high density region.

The proof of Theorem 1.3.1 is given in the appendix. It follows the idea in Horvath and Yandell (1988), who showed a stronger asymptotic result for the kernel based conditional empirical process with dimension $q = 1$. A corresponding result also holds for the conditional empirical process for $X_1 = 1$ and $X_2 = x_2$. Now define the decision rule of the test as

$$\text{“reject } \; H^1_{0,x_2} \; \text{ if } \; \hat{S}_1(x_2) > c_1\text{”}$$

where c_1 is the critical value of the first benchmark test that we will discuss in the following. The convergence results discussed above give the following proposition that characterizes the properties of the first one-sided nonparametric benchmark test.

Proposition 1.3.1. *Given Assumption 1.3.1-1.3.4 and that c_1 is a positive finite constant, we have:*

1. *If H^1_{0,x_2} is true,* $\lim_{n\to\infty} P(\textit{reject } H^1_{0,x_2}) \leq P\left(\sup_t \mathfrak{B}(t) > c_1\right) = \exp(-2c_1^2)$,

   with equality holds when equality in H^1_{0,x_2} holds.

2. *If H^1_{0,x_2} is false,* $\lim_{n\to\infty} P(\textit{reject } H^1_{0,x_2}) = 1$

The inequality in the first part of Proposition 1.3.1 implies that the test will never reject more often than $P(\sup_t \mathfrak{B}(t) > c_1)$ if the null hypothesis is true. The probability of rejection will be asymptotically equal to $P(\sup_z \mathfrak{B}(z) > c_1)$ when $F_1(.|x_2) = F_0(.|x_2)$ for the fixed value $x_2 \in \mathcal{W}$. The null is rejected with probability converging to 1 when it is not true. As is well known in the literature (see McFadden (1989) for instance) that $P(\sup_t \mathfrak{B}(t) > c_1) = \exp(-2c_1^2)$, the p-value of $\hat{S}_1$ is $\exp(-2(\hat{S}_1)^2)$. Then the null hypothesis is rejected if the p-value obtained is larger than α which is the nominal size of the test. The decision rule can also be stated in terms of critical values. Some standard critical values are 1.073 for the 10% significance level, 1.2239 for the 5% and 1.5174 for the 1%.

It is important to make clear that the convergence rate of our test statistic is slower than $\sqrt{n}$ and is dependent on the bandwidth because the Nadaraya Waston pointwise estimator we use to construct the test statistic relies more heavily on information in a small window around the focused value of the conditioning variable than information elsewhere. LW's test enjoys $\sqrt{n}$

convergence rate because they study whether the distributional partial effect has uniform sign for all values of the conditioning variable and hence construct the test statistic using all information from the sample set. However, we argue that although their convergence rate does not get worse when the dimension of the conditioning variable becomes larger, their test might be less useful in such cases as the null hypothesis becomes too strong to be interesting.

1.3.2 One-sided Nonparametric Testing: the Continuous Covariate Case

In this section, we propose tests for the null hypothesis $H^2_{0,x}$, or given a dataset whether the partial derivative of a continuous covariate on the conditional outcome distribution is uniformly negative. Assume the joint density of (Y, X) satisfies the following smoothness conditions. The conditions are slightly stronger than those in Section 1.3.1 since the test statistic to be proposed in this section is based on kernel derivative estimators instead of kernel regression estimators.

Assumption 1.3.5.

1. *Let $\mathcal{X}$ be the sets of X values that have positive densities. $\mathcal{X} \in \mathbb{R}^p$ is a compact set. X_1 is a scaler random variable and the first element in X.*

2. *The cumulative density function $G(y, x)$ be (absolutely) continuous. Let $s(y, x)$ be its corresponding probability distribution function.*

3. *Let $g(y, x) = \int 1(u \leq y)s(u, x)du$. $g(y, x)$ is uniformly bounded up until third derivatives.*

4. Let $f(x) = g(\infty, x)$. $f(x)$ is uniformly bounded up until third derivatives.

Given the definitions in above, the conditional distribution of the outcome variable $F_{Y|X=x}(y|x)$ is equal to the ratio of functions $g(y,x)$ and $f(x)$. The null hypothesis $H^2_{0,x}$ compares the partial derivative of $F_{Y|X=x}(y|x)$ with respect to x_1 with the zero function. Under the exogeneity Assumption A, we call the derivative as the DPE function of the continuous covariate X_1 on the outcome Y evaluated at $X = x$. Because for any small positive $\triangle h$ value, we have that

$$
\begin{aligned}
&\frac{1}{\triangle h}\left\{F_{Y|X_1=x_1+\triangle h, X_2=x_2}(y|x) - F_{Y|X_1=x_1, X_2=x_2}(y|x)\right\}\\
=&\frac{1}{\triangle h}\{E_Y[1(Y \leq y)|X_1 = x_1 + \triangle h, X_2 = x_2]\\
&- E_Y[1(Y \leq y)|X_1 = x_1, X_2 = x_2]\}\\
=&\frac{1}{\triangle h}\{E_e[1(H(x_1 + \triangle h, x_2, e) \leq y)|X_1 = x_1 + \triangle h, X_2 = x_2]\\
&- E_e[1(H(x_1, x_2, e) \leq y)|X_1 = x_1, X_2 = x_2]\}\\
=&\frac{1}{\triangle h}\{E_e[1(H(x_1 + \triangle h, x_2, e) \leq y)] - E_e[1(H(x_1, x_2, e) \leq y)]\}\\
=&\frac{1}{\triangle h}E_e[1(H(x_1 + \triangle h, x_2, e) \leq y) - 1(H(x_1, x_2, e) \leq y)], \forall y \in \mathcal{Y}
\end{aligned}
$$

If we relax the exogeneity assumption of X and e to the following selection under observables assumption, we could interpret the derivative function as the conditional distributional treatment effect of X_1 evaluated at $X_2 = x_2$.

Assumption C. $e \perp X_1 \mid X_2$.

Because for any positive small $\triangle h$ values,

$$
\begin{aligned}
&\frac{1}{\triangle h}\left\{F_{Y|X_1=x_1+\triangle h,X_2=x_2}(y|x) - F_{Y|X_1=x_1,X_2=x_2}(y|x)\right\}\\
=&\frac{1}{\triangle h}\{E_Y[1(Y \leq y)|X_1 = x_1 + \triangle h, X_2 = x_2]\\
&- E_Y[1(Y \leq y)|X_1 = x_1, X_2 = x_2]\}\\
=&\frac{1}{\triangle h}\{E_e[1(H(x_1 + \triangle h, x_2, e) \leq y)|X_1 = x_1 + \triangle h, X_2 = x_2]\\
&- E_e[1(H(x_1, x_2, e) \leq y)|X_1 = x_1, X_2 = x_2]\}\\
=&\frac{1}{\triangle h}\{E_e[1(H(x_1 + \triangle h, x_2, e) \leq y)|X_2 = x_2]\\
&- E_e[1(H(x_1, x_2, e) \leq y)|X_2 = x_2]\}\\
=&\frac{1}{\triangle h}E_e[1(H(x_1 + \triangle h, x_2, e) \leq y) - 1(H(x_1, x_2, e) \leq y)|X_2 = x_2], \forall y \in \mathcal{Y}.
\end{aligned}
$$

Assumption C is the counterpart of Assumption B with the selection variable being continuous.

Again the above exogeneity assumptions do not affect the asymptotic properties of the test statistic to be discussed. Examples where the test in $H^2_{0,x}$ is interesting include the effect of gas price (X_1) on the distribution of household new car MPG (Y) while controlling household characteristics (X_2) such as numbers of children, miles to work and so on, and the relationship between sons' wealth (X_1) and parents' wealth (Y) while controlling demographic characteristics (X_2) such as age and education and so on. In both examples, the covariate of interest (X_1) is continuous.

Use shorthand notion $F(y|x)$ for $F_{Y|X=x}(y|x)$ in the rest of the paper. For the test, first we define kernel estimators $\hat{f}(x)$, $\hat{F}(y|x)$ for $f(x)$ and $F(y|x)$

in the same way as in the last section. $\hat{f}^{(1)}(x)$, $\hat{g}^{(1)}(y, x)$ are kernel derivative estimators of partial derivatives of $f(x)$ and $g(y, x)$ with respect to x_1 defined as

$$\hat{f}^{(1)}(x) = -\frac{1}{nh^{p+1}} \sum_{i=1}^{n} K_1^{(1)} \left(\frac{X_i - x}{h} \right),$$
$$\hat{g}^{(1)}(y, x) = -\frac{1}{nh^{p+1}} \sum_{i=1}^{n} K_1^{(1)} \left(\frac{X_i - x}{h} \right) 1(Y_i \le y).$$

The kernel function K and the bandwidth h follow conditions in below.

Assumption 1.3.6. *Let K be a kernel function and $K_1^{(1)}$ its partial derivative with respect to the first argument. The kernel function satisfies:*

1. $\int K(\phi)d\phi = 1$, $\int \phi K(\phi)d\phi = 0$;
2. $\sup |K(\phi)| < \infty$, $\int |K(\phi)|d\phi < \infty$, $\int |\phi K(\phi)|d\phi < \infty$;
3. $\mu_2 = \int \phi^2 K(\phi)d\phi < \infty$
4. $\sup |K_1^{(1)}(\phi)| < \infty$, $\int |K_1^{(1)}(\phi)|d\phi < \infty$.

Assumption 1.3.7. *The bandwidth h is assumed to satisfy:*

1. $nh^{p+2} \to \infty$ *as* $n \to \infty$;
2. $nh^{p+6} \to 0$ *as* $n \to \infty$.

Denote $F^{(1)}(y|x)$ as the partial derivative of $F(y|x)$ with respect to x_1. Under Assumption 1.3.5-1.3.7, $F^{(1)}(y|x)$ could be consistently estimated by the Vinod and Ullah's (1988) kernel based estimator

$$\hat{F}^{(1)}(y|x) = \hat{f}(x)^{-1}\left[\hat{g}^{(1)}(y,x) - \hat{F}(y|x)\hat{f}^{(1)}(x)\right].$$

The convergence rate of Vinod and Ullah's (1988) estimator is $\sqrt{nh^{p+2}}$, for all $y \in \mathcal{Y}$. The bandwidth assumption in 1.3.7 is again stronger than that is necessary for consistency of the estimator as we want to eliminate some centering bias for the limiting distribution of $\hat{F}^{(1)}(.|x)$.

Given the definitions, we propose the test statistic to be:

$$\hat{S}_2(x) = A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}} \sup_{y\in\mathcal{Y}} \hat{F}^{(1)}(y|x),$$

where $A_2 = (\int K_1^{(1)}(\phi)^2 d\phi)^{-\frac{1}{2}}$. When X is single dimensional, $A_2 = (2/3)^{\frac{1}{2}}$ for the Epanechnikov kernel function and $(4\sqrt{\pi})^{\frac{1}{2}}$ for the Gaussian kernel function. For any fixed $x \in \mathcal{X}$, we know that $\hat{F}^{(1)}(.|x)$ is a random element in $(D[0,\bar{y}], \mathcal{B}(D[0,\bar{y}]))$. The following result characterizes its asymptotic behavior.

Theorem 1.3.2. *Under Assumption 1.3.1 and 1.3.5-1.3.7,*

$$A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\left(\hat{F}^{(1)}(.|x) - F^{(1)}(.|x)\right) \Rightarrow \mathfrak{B}(F(.|x)).$$

in $D([0,\bar{y}])$.

If we fix the y value, Theorem 1.3.2 degenerates to the standard weak convergence result of the kernel derivative estimator of $\frac{\partial E[1(Y\leq y)|X=x]}{x_1}$. The rate of convergence here is smaller than that in Theorem 1.3.1 because, for any fixed y, kernel derivative estimators converge slower than kernel regression estimators. The higher is the order of the derivative, the slower is the convergence rate. Now let c_2 be the critical value of the second benchmark test. We define decision rule of the test as

$$\text{“reject } H_{0,x}^2 \text{ if } \hat{S}_2(x) > c_2\text{”.}$$

From Theorem 1.3.2, we know that our second benchmark test shares the asymptotic properties with the first test including the consistency of the testing approach and the distribution free p-values and critical values.

Proposition 1.3.2. *Given Assumption 1.3.1 and 1.3.5-1.3.7 and that c_2 is a positive finite constant, we have:*

1. *If $H_{0,x}^2$ is true,* $\lim_{n\to\infty} P(\text{reject } H_{0,x}^2) \leq P\left(\sup_z \mathfrak{B}(z) > c_2\right) = \exp(-2c_2^2)$,
   with equality holds when equality in $H_{0,x}^2$ holds.
2. *If $H_{0,x}^2$ is true,* $\lim_{n\to\infty} P(\text{reject } H_{0,x}^2) = 1$.

1.4 Useful Extensions

1.4.1 Single-Index Tests for DPEs

The above two nonparametric benchmark tests are quite robust: only modest smoothness assumptions are required for the joint distribution of conditioning variables and the outcome. One cost for robustness is that convergence rates of the nonparametric test statistics decrease with the increase of the conditioning dimension. In this section, we consider semiparametric extensions for both benchmark tests, where the conditional distribution of the response variable is assumed to follow some single-index functional form. Or more specifically, assume that

$$F(y|x) = \tilde{F}(y|x\theta),$$

where θ is a p dimensional parameter vector lying in the parameter space Θ, for all $y \in \mathcal{Y}$. It is easy to see that $\tilde{F}(.|x\theta)$ also represents the conditional distributions of Y given $X\theta = x\theta$.

One set of conditions [2] that have economic interpretation and ensure the single index conditional distribution of the response variable are as follows.

Assumption 1.4.1. *Let θ is a p dimensional parameter vector in the compact parameter space Θ. Assume:*

1. *$Y = H(X\theta, e)$, where e is the unobserved determinant of the response Y,*

2. *$F_{e|X} = F_e$ or $F_{e|X} = F_{e|X\theta}$.*

The first part of Assumption 1.4.1 assumes that the response outcome depends only on the single index $X\theta$ and the unobserved determinant e. It is a popular structural assumption in the semiparametric identification literature. The second part assumes that the unobserved determinant e is either

[2] Another sufficient condition for the single-index assumption is that the cumulative distribution function of (Y, X) is equal to the cumulative distribution function of Y and $X\theta$.

independent of X or depends on X only through the single index $X\theta$. Let $\theta = (\theta_1\ \theta_2)$, where θ_1 is a scaler parameter. Normalize θ_1 to 1, then the DPE of a dummy X_1 variable under the single index assumption and exogeneity Assumption A is $\tilde{F}(.|1 + x_2\theta_2) - \tilde{F}(.|x_2\theta_2)$. The DPE function is $\frac{\partial}{\partial x_1}\tilde{F}(y|x\theta)$ if X_1 is continuous. When $\tilde{F}(y|x\theta)$ is non-monotonic for some $y \in \mathcal{Y}$, null hypotheses that we are interested in about whether X_1 has uniformly negative effect on the conditional outcome distribution become [3]:

$$H^3_{0,x_2} : \text{For fixed } x_2 \in \mathcal{W}, \tilde{F}(y|x_2\theta_2) \leq \tilde{F}(y|1 + x_2\theta_2) \quad \text{for all } y \in \mathcal{Y},$$
$$H^4_{0,x} : \text{For fixed } x \in \mathcal{X}, \frac{\partial}{\partial x_1}\tilde{F}(y|x\theta) \leq 0 \text{ for all } y \in \mathcal{Y}.$$

When the true value of θ is known, the above single-index null hypotheses could be easily tested by applying benchmark tests with the conditioning variable $X\theta$. When it is unknown, a two step testing approach could be employed where θ is estimated up-to-scale in the first step. Hall and Yao (2005) propose an iterative estimator that is $\sqrt{n}$-consistent and enables researchers to approximate conditional distributions using the single-index dimension reduction assumption. Meanwhile, given the fact that under Assumption 1.4.1 the conditional mean of the outcome variable $E(Y|X = x) = \int_{\mathcal{Y}} yd\tilde{F}(y|x\theta)$ also has a single-index representation, non-iterative (and hopefully easier to compute) estimators provided by single-index mean estimation literatures such as Powell, Stock, and Stoker's (1989) average partial derivative estimator and Horowitz and Hardle's (1996) estimator could also be used in the first step.

[3]When $\tilde{F}(y|x\theta)$ is monotonic for all $y \in \mathcal{Y}$, the distributional hypotheses for distributional partial effects could be answered by tests for the θ coefficient.

Assumption 1.4.2. *1. The estimator $\hat{\theta}$ is $\sqrt{n}$ consistent, that is for all θ in a compact space Θ, $\hat{\theta} - \theta = O_p(n^{-\frac{1}{2}})$,*

2. The kernel function K is M-th order differentiable.

Assumption 1.4.3. *The bandwidths are assumed to follow the additional condition that*

1. $n^M h_0^{2M+3} \to \infty$, $n^M h_1^{2M+3} \to \infty$.

2. $n^{M-1} h^{2M+1} \to \infty$.

The smoothness condition of the kernel function is used in Taylor expansions of the test statistics around the true θ value. And the additional bandwidth conditions ensure that the estimation error in estimating $\hat{\theta}$ does not affect the asymptotic distribution of the test statistic and so the statistic behaves, to first order, as if the value of θ is actually known.

Corollary 1.4.1. *Under Assumption 1.3.1-1.3.3 (1.3.1 and 1.3.5-1.3.6) with dimension $q = 1$ ($p = 1$) and the additional requirements in Assumption 1.4.1, 1.4.2 and 1.4.3, results in Proposition 1.3.1 (1.3.2) can be applied to the following test statistics respectively:*

$$\hat{S}_3(x_2) = A_1 \left[\frac{nh\hat{f}(1 + x_2\hat{\theta}_2)nh\hat{f}(x_2\hat{\theta}_2)}{nh\hat{f}(1 + x_2\hat{\theta}_2) + nh\hat{f}(x_2\hat{\theta}_2)} \right]^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \left(\hat{F}(y|1 + x_2\hat{\theta}_2) - \hat{F}(y|x_2\hat{\theta}_2) \right)$$

$$\left(\hat{S}_4(x) = A_2 \left[nh^3 \hat{f}(x\hat{\theta}) \right]^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \ \hat{F}^{(1)}(y|x\hat{\theta}) \right),$$

In Corollary 1.4.1, $\hat{f}$, $\hat{F}$ and $\hat{F}^{(1)}$ are the semiparametric density estimator, conditional distribution function estimator and derivative estimator for the conditional distribution function with respect to the covariate of interest. One advantage of the single index testing is that no sample splitting is involved in the semiparametric test statistics as long as the conditioning variable set includes one continuous element. Another is that no matter how many conditional variable are used by the researcher, the convergence rates of test statistics are always kept low, as if the tests were done with one single conditional variable. On the other hand, the single index tests described above have the disadvantage that the test results rely on the single index requirement in Assumption 1.4.1.

One thing to note is that under the single index setup, if econometricians or policy makers are interested in the causal influence of a covariate on the conditional distribution of the response but want to relax the exogeneity requirement, the igonorability condition in Assumption B and C could not be applied directly. Instead, we need the following counterparts of these two assumptions.

Assumption D. $(Y(0), Y(1)) \perp X_1 \mid X_2\theta_2$,

Assumption E. $e \perp X_1 \mid X_2\theta_2$.

Assumption D is for dummy X_1 covariates and Assumption E for continuous X_1 covariates. Neither assumptions and their nonparametric counterparts are nested within each other in general. But if in addition to Assumption

1.4.1, we also impose the assumption that $F(X_1|X_2) = \tilde{F}(X_1|X_2\theta)$, then Assumption D and E are implied by Assumptions B and C. [4]

In the case of discrete X_1, there is another dimension reduction situation other than Assumption 1.4.1 that might be of interest to empirical researchers.

Assumption 1.4.4. *Let θ_2 be a q dimensional parameter vector in the parameter space Θ_2. Assume:*

1. $Y = H(X_1, X_2\theta_2, e)$, *where e is the unobserved determinant of the response Y,*

2. $F_{e|X_1,X_2} = F_{e|X_1}$ *or* $F_{e|X_1,X_2} = F_{eX_1,X_2\theta_2}$,

Assumption 1.4.3 are weaker than Assumption 1.4.1. It ensures that $F(y|0,x_2) = F(y|0,x_2\theta_2) \overset{let}{\equiv} F_0(y|x_2\theta_2)$ and $F(y|1,x_2) = F(y|1,x_2\theta_2) \overset{let}{\equiv} F_1(y|x_2\theta_2)$. Then the null hypothesis of interest becomes

$$H_{0,x_2}^{3'} : \text{For fixed } x_2 \in \mathcal{W}, \tilde{F}_1(y|x_2\theta_2) \leq \tilde{F}_0(y|x_2\theta_2) \quad \text{for all } y \in \mathcal{Y}.$$

Let $\hat{\theta}_2$ be the $\sqrt{N}$-consistent estimator of θ_2, and $\hat{F}_0(y|x_2\hat{\theta}_2)$, $\hat{F}_1(y|x_2\hat{\theta}_2)$ be the nonparametric kernel based estimators of $\tilde{F}_0(y|x_2\theta_2)$ and $\tilde{F}_1(y|x_2\theta_2)$.

[4] Here we only show the statement for discrete covariate case. Suppose $P(X_1 = 1|X_2) = h(X_2\theta)$. Then Assumption D is implied by Assumption B because

$$\begin{aligned} P(X_1 = 1|e, X_2\theta) &= E[1(X_1 = 1)|e, X_2\theta_2] = E[E[1(X_1 = 1)|e, X_2]|e, X_2\theta_2] \\ &= E[E[1(X_1 = 1)|X_2]|e, X_2\theta_2] = E[h(X_2\theta)|e, X_2\theta_2] = h(X_2\theta) \\ &= P(X_1 = 1|X_2\theta_2). \end{aligned}$$

Define the test statistic as

$$\hat{S}_3'(x_2) = A_1 \left[\frac{n_1 h_1 \hat{f}_1(x_2\hat{\theta}_2) n_0 h_0 \hat{f}_0(x_2\hat{\theta}_2)}{n_1 h_1 \hat{f}_1(x_2\hat{\theta}_2) + n_0 h_0 \hat{f}_0(x_2\hat{\theta}_2)} \right]^{\frac{1}{2}} \sup_{y\in\mathcal{Y}} \left(\hat{F}_1(y|x_2\hat{\theta}_2) - \hat{F}_0(y|x_2\hat{\theta}_2) \right).$$

Parallel to Corollary 1.4.1, results in Proposition 1.3.1 could also be shown applicable to the test statistic $\hat{S}_3'(x_2)$, under the additional Assumption 1.4.2 and 1.4.4. The difference between this second single index testing method and the first one is that the second method is more robust as Assumption 1.4.4 is weaker than Assumption 1.4.1. Meanwhile when constructing the test statistic, the second method requires subsampling according to different values of the discrete X_1 covariate and therefore is less efficient.

1.4.2 Higher Order Stochastic Dominance Tests for DPEs

Sometimes first order stochastic dominance tests could be too restrictive if researchers care a lot about the part of distributional partial effect functions corresponding to the lower end of the distribution of the response variable. For example, policy makers focusing on inequality or poverty study might be interested in the stochastic dominance relationships of Lorenz curves or average poverty gaps from two income or wealth distributions. See Davidson and Duclos (2000) for uses of higher order stochastic dominance in inequality and poverty studies. Define $\mathcal{I}_j(y; h(t|x))$ as the function that integrates $h(t|x)$

to order $j-1$ with respect to t up to the value y:

$$\begin{aligned}
&\mathcal{I}_1(y;h(t|x)) = h(y|x),\\
&\mathcal{I}_2(y;h(t|x)) = \int_0^y h(t|x)dt = \int_0^y \mathcal{I}_1(y;h(t|x))dt,\\
&\mathcal{I}_3(y;h(t|x)) = \int_0^y \int_0^t h(s|x)dsdt = \int_0^y \mathcal{I}_2(y;h(t|x))dt,\\
&\ldots
\end{aligned}$$

Then depending on whether X_1 is discrete or continuous, the general hypotheses for testing stochastic dominance of order j could be written as:

$$\begin{aligned}
&H_{0,x_2}^{5.j} : \text{For fixed } x_2 \in \mathcal{W}, \mathcal{I}_j\left(y; F_1(t|x_2)\right) \leq \mathcal{I}_j\left(y; F_0(t|x_2)\right) \text{ for all } y \in \mathcal{Y},\\
&H_{0,x}^{6.j} : \text{For fixed } x \in \mathcal{X}, \mathcal{I}_j\left(y; \frac{\partial}{\partial x_1} F(t|x)\right) \leq 0 \text{ for all } y \in \mathcal{Y}.
\end{aligned}$$

When either of the above the null hypothesis is true and that both larger values and smaller variance for Y correspond to "better" outcomes, a marginal increase in X_1 has an unambiguously positive effect on outcomes in the distributional sense. Mathematically,

$$\begin{aligned}
\mathcal{I}_2\left(.; F_1(t|x_2)\right) \leq \mathcal{I}_2\left(.; F_0(t|x_2)\right) \Leftrightarrow &\int_{\mathcal{Y}} u(y)dF_1(y|x_2) \geq \int_{\mathcal{Y}} u(y)dF_0(y|x_2),\\
\mathcal{I}_2\left(.; \frac{\partial}{\partial x_1} F(t|x)\right) \leq 0 \Leftrightarrow &\frac{\partial}{\partial x_1} \int_{\mathcal{Y}} u(y)dF(y|x) \geq 0,\\
&\text{for all increasing and concave } u \text{ functions.}
\end{aligned}$$

Bawa (1975) has a detailed summary on properties of higher order stochastic dominance relationship. Define statistics for the higher order stochastic dominance tests as:

$$\hat{S}_5(x_2) = A_1 \left(\frac{n_1 h_1^q \hat{f}_1 n_0 h_0^q \hat{f}_0}{n_1 h_1^q \hat{f}_1 + n_0 h_0^q \hat{f}_0} \right)^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \left\{ \mathcal{I}_j \left(y; \hat{F}_1(t|x_2) \right) - \mathcal{I}_j \left(y; \hat{F}_0(t|x_2) \right) \right\},$$

$$\hat{S}_6(x) = A_2 \left[n h^{p+2} \hat{f} \right]^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \mathcal{I}_j \left(y; \hat{F}^{(1)}(t|x) \right),$$

where $\hat{f}_1$, $\hat{f}_0$, and $\hat{f}$ are shorthand notations $\hat{f}_1(x_2)$, $\hat{f}_0(x_2)$, and $\hat{f}(x)$ at fixed x_2 value. By Theorem 1.3.1, 1.3.2, the Continuous Mapping Theorem and the fact that the integral function $\mathcal{I}_j(.)$ is additive and continuous for all $j = 2, 3, ...$, the following corollary summarizes the asymptotic properties of the integrated processes used in the above test statistics.

Corollary 1.4.2.

$$A_1 \left[n_0 h_0^q \hat{f}_0(x_2) \right]^{\frac{1}{2}} \left\{ \mathcal{I}_j \left(y; \hat{F}_0(t|x_2) \right) - \mathcal{I}_j \left(y; F_0(t|x_2) \right) \right\} \Rightarrow \mathcal{I}_j \left(y; \mathfrak{B} \left(F_0(t|x_2) \right) \right),$$

$$A_1 \left[n_1 h_1^q \hat{f}_1(x_2) \right]^{\frac{1}{2}} \left\{ \mathcal{I}_j \left(y; \hat{F}_1(t|x_2) \right) - \mathcal{I}_j \left(y; F_1(t|x_2) \right) \right\} \Rightarrow \mathcal{I}_j \left(y; \mathfrak{B} \left(F_1(t|x_2) \right) \right),$$

$$A_2 \left[n h^{p+2} \hat{f}(x) \right]^{\frac{1}{2}} \left\{ \mathcal{I}_j \left(y; \hat{F}^{(1)}(t|x) \right) - \mathcal{I}_j \left(y; F^{(1)}(t|x) \right) \right\} \Rightarrow \mathcal{I}_j \left(y; \mathfrak{B}(F(t|x)) \right).$$

in $C([0, \bar{y}])$.

For these higher order stochastic dominance tests, asymptotic distributions of test statistics depend on the underlying conditional distributions of the response variable. That is, the p-values and critical values of the tests are no longer distribution free. In the following, we discuss a simulation approach for obtaining p-values for higher order stochastic dominance tests. It involves the use of artificial random numbers and the construction of the Brownian

Bridge process with random change of time to simulate a process that weakly converges to $\mathfrak{B} \circ F(.|x)$. Let $U_{i=1}^{n}$ denote a sequence of i.i.d. N(0, 1) random variables that are independent of the samples. It is well recognized (see Billingsley (1999) for instance) that the process

$$\mathfrak{B}^*(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1(i \leq nt) - t) U_i$$

weakly converges to a standard Brownian Bridge process $\mathfrak{B}$ on $D([0,1])$, $P(\mathfrak{B} \in C) = 1$. Since the estimator of the conditional distribution function $\hat{F}(.|x) \Rightarrow F(.|x)$ and is monotonically increasing from zero to one, we have that

$$\mathfrak{B}^* \circ \hat{F}(.|x) \Rightarrow \mathfrak{B} \circ F(.|x)$$

by Billingsley (1999). Use $\mathfrak{B}^*\left(\hat{F}(y|x)\right)$ to represent the process $\mathfrak{B}^* \circ \hat{F}(.|x)$ evaluated at the point $y \in \mathcal{Y}$, we have our simulated process

$$B^*\left(\hat{F}(y|x)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[1\left(i \leq n\hat{F}(y|x)\right) - \hat{F}(y|x)\right] U_i, \quad \forall y \in \mathcal{Y}.$$

The p-values of our higher order stochastic dominance tests for DPEs could be obtained from appropriate functionals of the simulated processes:

$$\hat{p}_{5.j} = P_U\left[\sup_{y \in \mathcal{Y}} \mathcal{I}_j\left(y; \mathfrak{B}^*\left(\hat{F}_0(u|x_2)\right)\right) > \hat{S}_5(x_2)\right],$$
$$\hat{p}_{6.j} = P_U\left[\sup_{y \in \mathcal{Y}} \mathcal{I}_j\left(y; \mathfrak{B}^*\left(\hat{F}(u|x)\right)\right) > \hat{S}_6(x)\right],$$

where $P_U(.)$ is the probability function associated with the normal random variable U_i's and is in practice equal to the fraction of simulations that have the interested supremum functional larger than the test statistic. The standard error of the simulated p-value is related to the number of simulations performed. See Barrett and Donald (2002) or Hansen (1996) for discussions on simulation number and the standard error of the simulated p-value. The following results describe the asymptotic characteristics of higher order stochastic dominance tests and simulated p-values.

Proposition 1.4.1. *Given Assumption 1.3.1-1.3.4 (1.3.1, 1.3.5-1.3.7) and assuming that $\alpha < \frac{1}{2}$, a test for j-th order stochastic dominance test for discrete (continuous) distributional partial effects based on the rule*

$$\text{“}\textit{reject } H^{5.j}_{0,x_2}(H^{6.j}_{0,x}) \textit{ if } \hat{p}_{5.j}(\hat{p}_{6.j}) < \alpha\text{”}$$

satisfies the following asymptotic properties:

1. *If $H^{5.j}_{0,x_2}(H^{6.j}_{0,x})$ is true, $\lim_{n\to\infty} P\left[\textit{reject } H^{5.j}_{0,x_2}(H^{6.j}_{0,x})\right] \leq \alpha$, with equality holds when equality in $H^{5.j}_{0,x_2}(H^{6.j}_{0,x})$ holds.*

2. *If $H^{5.j}_{0,x_2}(H^{6.j}_{0,x})$ is false, $\lim_{n\to\infty} P\left[\textit{reject } H^{5.j}_{0,x_2}(H^{6.j}_{0,x})\right] = 1$.*

Given the weak convergence property of the simulated process for changed time Brownian Bridge, proof for the proposition is the same as that in Barrett and Donald (2003) and is hence omitted.

1.4.3 Two-sided Tests for DPEs

In this section, we discuss nonparametric two-sided tests corresponding to the one-sided nonparametric benchmark tests. Two-sided single-index tests could be constructed similarly. Consider the following two-sided null hypotheses:

$$
\begin{aligned}
&H^7_{0,x_2} : \text{For fixed } x_2 \in \mathcal{W}, F_1(y|x_2) = F_0(y|x_2) \text{ for all } y \in \mathcal{Y}, \text{ and} \\
&H^8_{0,x} : \text{For fixed } x \in \mathcal{X}, \frac{\partial}{\partial x_1} F(y|x) = 0 \text{ for all } y \in \mathcal{Y}.
\end{aligned}
$$

Both null hypotheses imply that the DPE function of X_1 evaluated at $X_2 = x_2$ ($X = x$) is the zero function, or X_1 has no effect on the conditional distribution of the response variable at all. We define the test statistics to be

$$
\begin{aligned}
\hat{S}_7(x_2) &= A_1 \left(\frac{n_1 h_1^q \hat{f}_1(x_2) n_0 h_0^q \hat{f}_0(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)} \right)^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \left| \hat{F}_1(y|x_2) - \hat{F}_0(y|x_2) \right|, \\
\hat{S}_8(x) &= A_2 \left[n h^{p+2} \hat{f}(x) \right]^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \left| \hat{F}^{(1)}(y|x) \right|,
\end{aligned}
$$

where the supremum in $\hat{S}_1(x_2)$ and $\hat{S}_2(x)$ for the one-sided benchmark tests are replaced with the supremum of absolute values. By the weak convergence results in Theorem 1.3.1, 1.3.2 and the decision rule

$$
\text{“reject } H^7_{0,x_2}(H^8_{0,x}) \text{ if } \hat{p}_7(\hat{p}_8) < c_7(c_8)\text{”},
$$

the asymptotic characteristics of the two-sided tests could be derived as follows.

Proposition 1.4.2. *Given Assumption 1.3.1-1.3.4 (1.3.1 and 1.3.5-1.3.7) and that $c_7(c_8)$ is a positive finite constant, we have:*

1. *If $H^7_{0,x_2}(H^8_{0,x})$ is true, $\lim_{n\to\infty} P[\text{reject } H^7_{0,x_2}(H^8_{0,x})] = P\left[\sup_t |\mathfrak{B}(t)| > c_7(c_8)\right] = 1 - \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2 c_7^2(c_8^2))$.*

2. If $H^7_{0,x_2}(H^8_{0,x})$ is false, $\lim_{n\to\infty} P[$*reject* $H^7_{0,x_2}(H^8_{0,x})] = 1$.

Press, Teukolsky, Vetterling and Flannery (1996) provides programs for calculating the probability function stated above. Some standard critical values are 1.2238 for the 10% significance level, 1.3581 for the 5% and 1.6276 for the 1%.

1.5 Monte Carlo Results

In this section we conduct a number of Monte Carlo experiments to examine how asymptotic properties of benchmark and extension tests discussed in Sections 1.3 and 1.4 hold in small datasets. The first set of experiments focus on small sample behaviors tests (one-sided or two-sided, nonparametric or single index, first order or second order stochastic dominance) studying the effect a discrete covariate on the conditional outcome distribution. For each experiment, independent variable X_1 is randomly drawn to take value 0 and 1 each with probability one-half while X_2, $X_3 \sim N(0, 0.3^2)$. To evaluate the test performance when the null hypothesis is true and equality in the null is satisfied, the response variable Y is generated independently from $N(0, 0.3^2)$. The DPE of X_1 is hence equal to the zero function for any possible X_2, X_3 values. We test the null that the DPE of X_1 is uniformly negative with the effect evaluated at $(X_2, X_3) = (0, 0)$. To see test performance when the null hypothesis is false, Y is generated following a simple random coefficient model $Y = X_1 b - X_1(X_2 + X_3)$, where $b \sim U[0, 1]$ is a random coefficient. The DPE is NOT uniformly negative when $X_2 + X_3 > 0$. We test the null that the DPE of X_1 is uniformly negative with the effect evaluated at $(X_2, X_3) = (0.2, 0.2)$.

For each of the two data generating processes, we consider three sets of conditioning variables in testing: the single-dimensional conditioning variable $SI = X_2 + X_3$, the two-dimensional conditioning variable set (X_2, X_3) and the single-index conditioning variable $\hat{SI} = X_2 + X_3\hat{b}$, where $\hat{b}$ is Powell Stock and Stoker's (1989) up-to-scale weighted average derivative estimator obtained under the single index Assumption 1.4.3. Notice that the weighted average derivative estimator requires higher order kernel density function. In this section, we follow Horowitz and Wolfgang (1996) and use the 4th order kernel function $K(t) = (105/64)(1 - 5t^2 + 7t^4 - 3t^6)1(|t| \leq 1)$ and bandwidth $5n^{-1/6}\sigma$ for the weighted average derivative estimation, where σ is the standard deviation of the conditioning variable. The Epanechnikov kernel is used for the test and bandwidth are chosen to be $cn_0^{-\frac{1}{4.75}}\sigma$, $cn_1^{-\frac{1}{4.75}}\sigma$ for the first and third tests with single-dimensional conditioning variable and $cn_0^{-\frac{1}{5.75}}\sigma$, $cn_1^{-\frac{1}{5.75}}\sigma$ for the second case. The constant c is allowed to take values 2, 2.5 or 3 to see whether the performance of tests is sensitive to the bandwidth choice.

The first part of Table 1.1 reports rejection probabilities of each one-sided test using 5% significance level and 5000 simulations. I find that all nonparametric and semiparametric tests report rejection probabilities close to 5% when the null is true and equality in the null held, and rejection probabilities going to 1 with sample size when the null is false. Therefore, we conclude that all three one-sided nonparametric semiparametric DPE tests have good small sample behavior and their performances are not very sensitive to bandwidth choices. Moreover, by comparing performance of the first two nonparametric tests we notice that when the null is false, rejection probabilities go to 1 slower when we have more conditioning variables in tests. Comparing performance of the single-index test with the two nonparametric test we notice that the semiparametric coefficient estimator, as is predicted by the theory, does not affect the test performance much in small samples. The rejection proportion of the single-index test converges to 1 significantly faster than its nonparametric counterpart with two-dimensional conditioning variables.

Secondly, we use the same data generating processes to study small sample behaviors of two-sided nonparametric and semiparametric tests and one-sided second order stochastic dominance tests. We see from the second and third section of table 1.1 that all of these extended tests also have good small sample behaviors: when the null is correct and equality in null held, the rejection probabilities are close to 5%; when the null hypothesis is false, the rejection rate converges to 1. The slower convergence rate of tests with more conditioning variables and the advantage of using single-index testing (when the single index assumption is valid) are the same as discussed.

Then we study the small sample behavior of our second benchmark test concerning DPEs of continuous covariates. Due to the space limit, we only report the results from one-sided nonparametric tests with single-dimensional conditioning variable. [5]We also compare the test performance to that of LLW

[5]Small sample behaviors of two-sided tests and second-order stochastic dominance tests are as good.

Table 1.1: Rejection Proportions of DPE Tests: Discrete Covariate Case

	Nonparametric 1			Nonparametric 2			Semiparametric		
Sample Size	c=2	c=2.5	c=3	c=2	c=2.5	c=3	c=2	c=2.5	c=3
				One-sided First Order Tests					
When the null is true and equality in null held:									
N=100	0.040	0.046	0.050	0.039	0.047	0.052	0.043	0.046	0.050
N=250	0.043	0.051	0.051	0.044	0.051	0.056	0.045	0.046	0.051
N=500	0.046	0.048	0.048	0.046	0.049	0.051	0.044	0.049	0.048
When the null is not true:									
N=100	0.193	0.185	0.170	0.105	0.122	0.132	0.188	0.180	0.167
N=250	0.755	0.773	0.770	0.403	0.507	0.576	0.741	0.762	0.762
N=500	0.987	0.991	0.994	0.792	0.890	0.944	0.985	0.991	0.993
				Two-sided First Order Tests					
When the null is true and equality in null held:									
N=50	0.041	0.041	0.045	0.040	0.046	0.051	0.040	0.042	0.045
N=100	0.040	0.042	0.045	0.041	0.051	0.060	0.040	0.045	0.047
N=250	0.050	0.050	0.048	0.040	0.046	0.049	0.050	0.048	0.049
When the null is false:									
N=50	0.749	0.893	0.968	0.534	0.820	0.962	0.786	0.913	0.974
N=100	0.989	0.998	1.000	0.900	0.995	1.000	0.990	0.999	1.000
N=250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				One-sided Second Order Tests					
When the null is true and equality in null held:									
N=100	0.064	0.062	0.064	0.061	0.063	0.067	0.060	0.061	0.062
N=250	0.052	0.052	0.059	0.058	0.052	0.052	0.053	0.054	0.055
N=500	0.047	0.053	0.052	0.053	0.054	0.054	0.050	0.049	0.052
When the null is false:									
N=100	0.650	0.680	0.669	0.586	0.630	0.630	0.647	0.681	0.666
N=250	0.937	0.953	0.957	0.850	0.898	0.924	0.935	0.951	0.958
N=500	1.000	1.000	1.000	0.987	0.997	0.999	1.000	1.000	1.000

using the same DGP although our second benchmark test and LLW's test are based on different null hypotheses. LLW's test, as discussed earlier, has stronger null hypothesis and much slower computation. Simulation is done 5000 times for our test but only 200 times for LLW's due to their computational speed constraint. This difference in simulation numbers is not expected to affect the validity of small sample performance comparison.

Consider the DGP with $X \sim U[0,1]$. Firstly, the response variable is independently generated following $N(0, 0.3^2)$. We test whether the DPE of X is uniformly negative at $X = 0.5$ for our test, and whether it is uniformly negative at all X values for LLW's test. Since null hypotheses for both tests are true and equality in the nulls held, both tests shall reject the null with probability equaling to the significance level. Then we generate the response variable through the relationship $Y = Xb + 2(X - 0.25)^2$, where $b \sim U[0,1]$ is a random coefficient, and test small sample performances of both tests when null hypotheses are false. DPE of X on the response Y is NOT uniformly negative when $X < 0.25$. We test whether the DPE of X is uniformly negative at $X = 0.1$ for our test and whether it is uniformly negative for all X values for LLW's test. Rejection proportions for both tests shall go to 1 when the sample size increases.

Table 1.2 reports rejection proportions of the above Monte Carlo experiments using 5% significance level. Bandwidths are chosen to be $h = cn^{-\frac{1}{6.75}}$ for all experiments with c=1, 1.25 or 1.5. We see that our tests perform well: the rejection proportions are close to 5% when the null is true and go to 1

Table 1.2: Rejection Proportions of DPE Tests: Discrete Covariate Case

	Our Test			LLW1*			LLW2*		
Sample Size	c=1	c=1.25	c=1.5	c=1	c=1.25	c=1.5	c=1	c=1.25	c=1.5
One-sided Nonparametric First Order Tests									
When the null is true and equality in null held:									
N=250	0.039	0.037	0.040	0.235	0.155	0.070	0.005	0.025	0.010
N=500	0.038	0.042	0.041	0.295	0.190	0.080	0.005	0.010	0.005
N=1000	0.041	0.046	0.037	0.345	0.280	0.155	0.000	0.000	0.000
When the null is not true:									
N=250	0.371	0.201	0.114	0.190	0.170	0.165	0.130	0.125	0.135
N=500	0.860	0.661	0.464	0.420	0.585	0.595	0.440	0.630	0.620
N=1000	0.998	0.990	0.963	0.950	0.990	0.990	0.950	0.990	0.995

Note: *LLW1 calculates the rejection proportions using the asymptotic expansion approach while LLW2 calculates the rejection proportions using asymptotic distribution approach. See LLW for details.

quickly as the sample sizes increase when the null is false; the rejection proportions are not very sensitive to bandwidth choices. The test performance of LLW's test is almost as good as ours when the null is false but is not very stable when the null is true and equality in null held.

In conclusion, our nonparametric and semiparametric tests on DPEs have good small sample behaviors. For nonparametric test, larger sample size is needed for tests with more conditioning variables. But if econometricians or policy makers are willing to assume that the conditional cumulative distribution of the response variable is of some single-index form, they could use semiparametric techniques to reduce the dimension of their conditioning variable and test more efficiently.

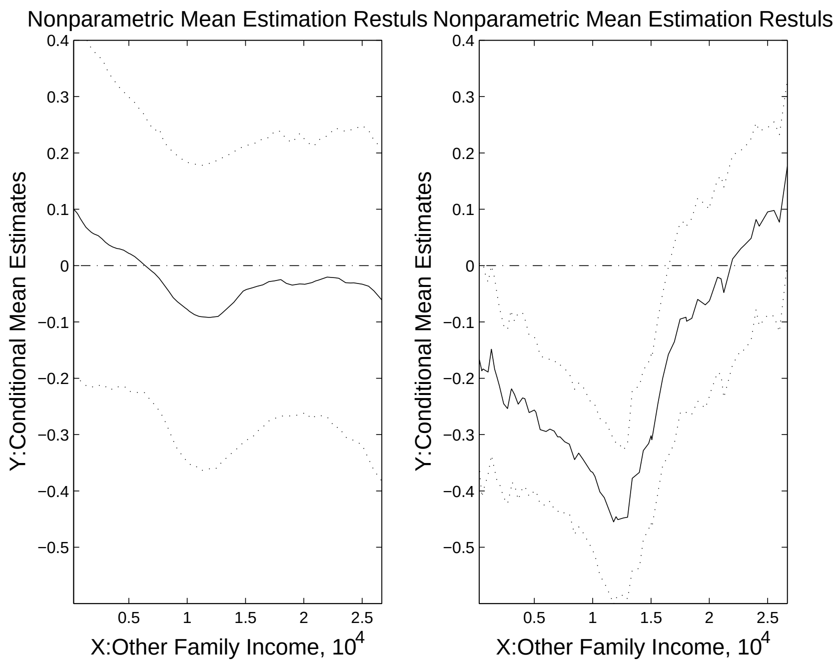
1.6 Empirical Example

In this section, we present an empirical example concerning children and father's labor supply. We study the effect of children gender on father's labor income distribution and the association between other family income and father's labor income distribution. The data we use comes from Angrist and Evans (1998), which is from the 5% Public Use Microdata Sample (PUMS) of 1990 Census. Our sample is limited to white households with working parents and 2 same sex children less than 10 years old. Fathers are 36 years old at the time of the survey. I didn't group fathers with different ages because when studying the association between other family income and fathers' labor income I want to rule out the positive relationship coming from the positive relationship between father and mothers' age. [6] The subtracted subsample includes 2408 observations.

Let X be other family income, Z be a dummy equals to one when both children are boys and Y be father's labor income. First we report in Figure 1.1.a PE estimates of children gender conditioning on different other family income values and their bootstrapped 95% confidence intervals. Lundberg and Rose (2002) found that fathers' wage rates increase more in response to the births of sons than to the births of daughters. However, we do not observe

[6] Observations with allocated child gender, child and parents age and parents income are excluded following Angrist and Evans (1998). Families with 2 same sex children are chosen for the study of children gender and father's labor income because if there is any effect of children gender, having 2 sons compared to 2 daughters is believed to exaggerate the effect in scale and is easier to detect.

Figure 1.1: PE Estimates of Children Gender and Other Family Income



Note: The kernel used is Epanechnikov. The bandwidth used is $2.34N^{-1/5}\sigma$ for the left graph and $2.15N^{-1/7}\sigma$ for the right graph. Mean estimations are performed conditioning on other family income values taking from quantile 0.01, 0.02 to 0.80. Other family income values lager than 80% quantile are not used for testing due to the low density. (The other family income distribution is highly right skewed.)

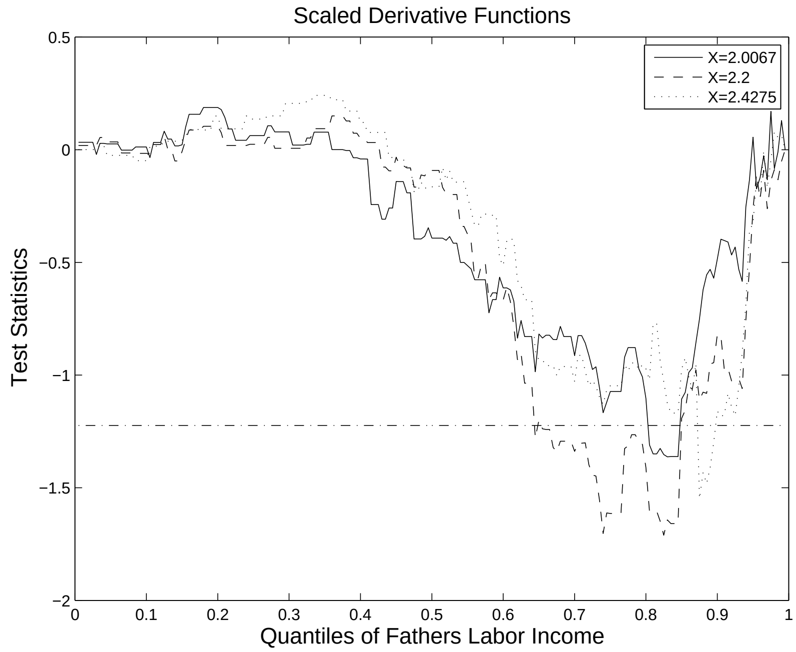
Figure 1.2: DPE Tests for Children Gender and Other Family Income

Note: The kernel used is Epanechnikov. The bandwidth used is $2.5N^{-1/4.75}$ for the left graph and $1.25N^{-1/6.75}$ for the right graph. DPE tests are performed conditioning on other family income values taking from quantile 0.01, 0.02 to 0.80. Other family income values lager than 80% quantile are not used for testing due to the low density. (The other family income distribution is highly right skewed.)

that fathers with two sons on average earn more: children gender is found irrelevant to fathers' average labor income in our sample. Test statistics for uniformly zero DPE function of children gender are reported in Figure 1.2.a. We cannot reject the null that children gender has no effect on father's labor income with 5% significance level at all values that we used of other family income.

Then we study the relationship between other family income and fathers' labor income. Figure 1.1.b shows that the relationship between other family income and average fathers' labor income is not monotonic along the support of other family income. At small values of other family income, an increase in other family income is associated with lower average fathers' labor income, which is consistent with the negative income effect story. At large values of other family income, the association is not significant. Then we perform our benchmark distributional tests holding levels of other family income fixed at different values. We see from Figure 1.2.b that we reject for small other family income the null that a marginal increase is associated with "better" fathers' labor income distribution and cannot reject the null that an increase is associated with "worse" father's labor income distribution. The result is again consistent with the negative income effect theory. For larger other family income, we reject the null that a marginal increase is associated with "worse" fathers' income distribution and cannot reject the null that an increase is associated with "better" fathers' income distribution. This means that there exist some positive association between other family income and fathers' labor income. And the association, like positive marriage sorting, dominates the negative income effect at least for some part of fathers' labor income distribution. Actually, we can also roughly tell where in the father's labor income conditional distribution we have this dominating positive association by looking at the estimated derivative functions of the conditional fathers' labor income distribution. But more rigorous set and confidence re-

Figure 1.3: Rescaled DPE Functions for Other Family Income



Note: The kernel used is Epanechnikov. DPE tests are performed conditioning on other family income values at quantile 0.65, 0.70 to 0.75. The bandwidth used is $1.25N^{-1/6.75}$.

gion estimation is out of the scope of this paper. It is a topic worthy of for future study.

We report in Figure 1.3 the scaled estimated derivative functions of the conditional father's labor income distribution with respective to other family income, conditional on three large values of other family income (at 65%, 70% and 75% quantiles). We find that for all three estimated functions, positive associations between other family income and fathers' labor income are detected only at relatively high quantiles of fathers' labor income. Then why can't we just do nonparametric quantile regression at the, say, 75% quantile of fathers' labor income? By doing so, we can still reject the null hypothesis of negative association at the 75% quantile. A lot empirical researchers who are familiar with quantile regression techniques might ask this question. But note, in this case, we are not interested in the result at the 75% quantile of fathers' labor income per se – we will happily report any rejection of the negative association, if found, at any other outcome quantiles, regardless its 50%, 60% or 75%. So what we actually want to know is whether there is a positive association dominating the negative income effect at any part of the fathers' labor income distribution. We cannot use test results obtained at specific outcome quantiles to reach this conclusion. There are so many quantiles that we could test on. We could actually have too many chances to reject the null! Therefore, we will need our distributional partial effect tests as long as we are not interested in results at specific outcome quantiles but results for the whole outcome distribution.

1.7 Conclusion

In this paper we have proposed Kolmogorov-Smirnov type tests for testing whether the distributional partial effect function of a covariate, evaluated at fixed values of itself and other conditioning variables, is uniformly positive, negative or zero. The tests are easy to implement. Critical values or p-values, except for higher order stochastic dominance extensions, are distributional free and hence involve no simulation or bootstrapping methods. We have also shown that the tests perform well in finite samples. A few related topics of the distributional partial effects discussed in this paper are worthy of future research. First of all, it is interesting to construct confidence regions where the distributional partial effect function of a covariate satisfies some moment conditions, such as uniform negativity. The related set estimation and confidence region literatures include Chernozhukov, Hong, and Tamer (2007) and Andrews and Shi (2010), among others. Our DPE tests could be potentially improved in power using methods suggested in Linton, Song and Whangs (2010) and Donald and Hsu (2010). The idea of comparing conditional distributions in the stochastic dominance sense could be useful under the regression discontinuity context, too. But the nonparametric estimators concerned are better modified to utilize local linear regression techniques so as to deal with the boundary problem involved in the regression discontinuity setup. Last but not least, DPE tests allowing for data dependence are worthy of future study as the stochastic dominance concepts have been widely applied in the finance literature. Linton, Maasoumi and Whang (2005) provides a related work.

Chapter 2

Estimation of Censored Panel-data Models with Slope Heterogeneity[1]

2.1 Introduction

The use of panel data to control for individual heterogeneity is pervasive in economics. Most often, the approach taken in panel-data models—whether they are random-effects or fixed-effects models—is to assume that the unobserved heterogeneity enters additively. In linear models, this form of additive heterogeneity allows the overall level of the outcome to vary with the heterogeneity but restricts the marginal effects of covariates to be the same within the population (invariant to the heterogeneity). Generalization of the linear additive fixed-effects model to fixed effects in slopes has been considered by Cornwell, Schmidt, and Sickles (1990) and Polachek and Kim (1994), and the estimation approach for such models is described in Wooldridge (2002).[2] The literature on heterogeneous slopes in nonlinear panel-data models is compara-

[1]This chapter borrows extensively from an earlier joint work with Jason Abrevaya (2010).

[2]Wooldridge (2005) provides conditions under which standard fixed-effects estimators consistently estimate the population average of the fixed-effects slopes.

tively sparse. For binary-choice models, Thomas (2006) considers an extension of the fixed-effects logit model that allows for heterogeneous linear trends. The approach relies on the exponential form of the logit model and, therefore, does not provide a general estimation strategy for non-linear models.

In this paper, we consider estimation of a censored panel-data model with heterogeneous slopes. Honoré (1992) provided estimators of the censored panel-data model with additive fixed effects. These estimators are based upon a trimming strategy that utilizes the stationarity of the composite error term (idiosyncratic disturbance plus fixed effect), a strategy that does not obviously generalize to the case of fixed-effects in slopes. Instead, we employ the projection approach of Chamberlain (1984) and Mundlak (1978) that is frequently used in non-linear panel-data applications. Specifically, we model the heterogeneous slopes as being related to the covariates and also having a random component to them (*correlated random slopes* models). We also consider heterogenous-slopes models in which there is no relationship between the slopes and covariates (*random slopes* models). We provide maximum likelihood estimation (MLE) and censored least-absolute deviations (CLAD) estimation methods for these models, the former based upon a fully parametric specification of the random components and the latter based upon a conditional median assumption on the composite error term of the model.

An outline of this paper is as follows. Section 2 formally introduces the censored panel-data model with individual-specific slope heterogeneity. Focusing upon the case of normally distributed disturbances (yielding

Tobit-type models), the random slopes model and correlated random slopes model are considered separately. The MLE estimator, in the context of a Chamberlain- or Mundlak-type model for the heterogeneous slopes, is described in detail in Section 2.1. Specification tests are provided to test the slope-heterogeneity models against nested alternatives. As in standard censored regression models, it is useful to make a distinction between censoring that arises due to data-coding issues (e.g., bottom-coding or top-coding) and censoring that arises due to corner-solution outcomes (a distinction that is explained at length in Wooldridge (2002)). For corner-solution outcomes, we are primarily interested in partial effects on the observed dependent variable (rather than the underlying latent dependent variable used in the model). We provide details for estimation of average partial effects in the case of corner-solution outcomes. Section 2.2 relaxes the normality assumption and proposes a CLAD estimator under a conditional median assumption. Section 3 applies the proposed estimators to an empirical study of Dutch household portfolio choice. We find strong evidence of correlated random slopes for the age variables, indicating that the age profile of portfolio adjustment varies significantly with other household characteristics.

2.2 Models and Estimators

2.2.1 Maximum Likelihood Estimation

Consider the following censored panel-data model with individual-specific slope heterogeneity, where the dependent variable is assumed to be left-censored at zero, without loss of generality:

$$
\begin{aligned}
y_{it}^* &= X_{it}\beta + Z_{it}c_i + u_{it} \quad (i = 1, 2, ..., N;\ t = 1, 2, ..., T) \qquad (2.1) \\
y_{it} &= \max\{0, y_{it}^*\}
\end{aligned}
$$

The observed data is $\{(y_{it}, X_{it}, Z_{it}) :\ i = 1, \ldots, N;\ t = 1, \ldots, T\}$, where X_{it} is a $1 \times p_1$ covariate vector and Z_{it} is a $1 \times p_2$ covariate vector. Slope homogeneity is assumed for the X variables, but the Z variables have slope heterogeneity (c_i varying with i). For individual components of Z_{it}, the following notation is used:

$$Z_{it}c_i = z_{1it}c_{1i} + z_{2it}c_{2i} + \cdots + z_{p_2it}c_{p_2i}.$$

For expositional purposes, the panel is assumed to be balanced (T observations for each i), but our approach can be easily extended to the case of unbalanced panels.

The following strict exogeneity assumption is made:

Assumption 1. (Strict exogeneity) u_i *is independent of* (X_i, Z_i, c_i),

where $u_i \equiv (u_{i1}, \ldots, u_{iT})'$. Assumption 1 is stronger than the strict exogeneity assumption traditionally made for linear models (where only conditional-mean independence is assumed) but is commonly made for non-linear models. Note that Assumption 1 does not rule out serial correlation in the u_{it}'s, but the presence of a lagged dependent variables ($y_{i,t-1}$) as a covariate within X_{it} or Z_{it} would violate Assumption 1.

Note that the heterogeneous-intercept model is a special case of model (2.1) in which Z_{it} is a constant. If Z_{it} contains a constant and a time variable (that is, $Z_{it} = (1, t)$), then (2.1) becomes a heterogeneous-trend censored panel-data model. More generally, model (2.1) allows heterogeneity to enter through the c_i coefficients on the Z_{it} covariates. In the following subsections, we consider alternative assumptions concerning the relationship between c_i and (X_i, Z_i).

Random-Slopes Tobit Models

Consider the case in which the slope heterogeneity is random, specifically:

Assumption 2. (Random slopes) c_i *is independent of* (X_i, Z_i).

Letting $\lambda \equiv E(c_i) = E(c_i|X_i, Z_i)$, the "residual" $a_i \equiv c_i - \lambda$ is also independent of (X_i, Z_i). The latent-variable model in (2.1) can then be rewritten

$$y_{it}^* = X_{it}\beta + Z_{it}\lambda + Z_{it}a_i + u_{it}. \tag{2.2}$$

In order to estimate the resulting model with likelihood methods, a parametric assumption on the conditional distributions of u_{it} and a_i is made. We assume normality here, primarily for convenience, although other distributional assumptions could be used.

Assumption 3. (Normality) *(i)* $u_{it}|X_i, Z_i, a_i \sim N(0, \sigma_u^2)$ *for* $t =, 1 \ldots, T$*; (ii)* $a_i|X_i, Z_i \sim N(0, \Sigma)$.

Assumption 3 restricts the variance of u_{it} to be constant over t, but this restriction can be relaxed at the expense of introducing additional parameters. No assumption on the serial correlation between u_{it}'s is made here. The Σ covariance matrix is $p_2 \times p_2$. Combining the two parts of Assumption 3 yields

$$(Z_{it}a_i + u_{it})|X_i, Z_i \sim N(0, \sigma_u^2 + Z_{it}\Sigma Z_{it}'). \tag{2.3}$$

As a result, model (2.2) can be viewed as a pooled Tobit model with heteroskedasticity, where the form of heteroskedasticity is given by (2.3).

Denote $\theta \equiv [\beta \quad \lambda]$, so that the parameters of interest are $(\theta, \sigma_u^2, \Sigma)$. Further, let $M_{it} \equiv [X_{it} \quad Z_{it}]$ and $D_{it} \equiv 1\{y_{it} > 0\}$. Under Assumptions 1–3, a consistent estimator of these parameters (which we call the *random-slopes Tobit estimator*) is obtained by maximizing the following partial log likelihood function:

$$\begin{aligned} LL(\tilde{\theta}, \tilde{\sigma}_u^2, \widetilde{\Sigma}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Bigg[&(1 - D_{it})\log\left(1 - \Phi\left(\frac{M_{it}\tilde{\theta}}{(\tilde{\sigma}_u^2 + Z_{it}\widetilde{\Sigma}Z_{it}')^{\frac{1}{2}}}\right)\right) \\ &+ D_{it}\log\frac{1}{(\tilde{\sigma}_u^2 + Z_{it}\widetilde{\Sigma}Z_{it}')^{\frac{1}{2}}}\phi\left(\frac{y_{it} - M_{it}\tilde{\theta}}{(\tilde{\sigma}_u^2 + Z_{it}\widetilde{\Sigma}Z_{it}')^{\frac{1}{2}}}\right)\Bigg], \end{aligned} \tag{2.4}$$

subject to the constraint that $\widetilde{\Sigma}$ is a positive semi-definite symmetric matrix.

Several points are worth mentioning with respect to the pooled Tobit estimator above. When Z_{it} includes a constant term, σ_u^2 is not separately identified from the variance parameter in Σ corresponding to the constant term. If Z_{it} includes a set of categorical variables (with mutually exclusive categories), the covariance parameters within Σ corresponding to these components are not identified; in terms of the likelihood function, these parameters get multiplied by the product of category dummies (which will always be zero). As a practical matter, we can just set these covariance parameters to be equal to zero in estimation. One may also impose the restriction that Σ is a diagonal matrix (i.e., no covariances between the elements of a_i), which reduces the number of parameters to be estimated and can also greatly reduce computation time.

When each component of a_i has a non-zero variance (i.e., non-zero diagonal elements for Σ), the partial likelihood estimator has a standard asymptotic distribution — standard errors can be calculated using the robust sandwich formula, as in Wooldridge (2002). If one or more components of a_i have zero variance, then inference may be complicated by the fact that variance parameters lie on the boundary of the parameter space. As Andrews (1999, 2001) points out, the asymptotic sandwich formula and also the standard bootstrap method are generally invalid in this situation. One valid approach discussed by Andrews (1999, 2001), which we employ in our empirical application, is to use bootstrap re-sampling based upon subsamples smaller than the original sample size N.[3]

If, in addition to Assumptions 2 and 3, one is willing to make an assumption about the serial correlation pattern in the u_{it}'s (or, as a special case, assume that there is no serial correlation), a full-information maximum like-

[3]This approach is sometimes called the *m-out-of-N* bootstrap and utilizes sampling with replacement. Alternatively, one could do subsampling where the sampling is done without replacement. In either case, the standard errors must be re-scaled appropriately based upon m and N.

lihood estimator can be used. If $u_i|(X_i, Z_i, a_i) \sim N(0, \Sigma_u)$, the log-likelihood function is given by

$$LL(\tilde{\theta}, \widetilde{\Sigma}_u, \widetilde{\Sigma}) = \frac{1}{N}\sum_{i=1}^{N} log\left[\int_{-\infty}^{\infty} f(y_1, ..., y_T|M_i, a; \widetilde{\Sigma}_u, \tilde{\theta})f(a; \widetilde{\Sigma})\mathrm{d}a\right], \quad (2.5)$$

where $f(y_1, ..., y_T|M_i, a; \Sigma_u, \beta, \lambda)$ is the joint distribution of $(y_1, \ldots, y_T)$ conditional on the covariates and the slopes, and $f(a; \Sigma)$ is the multivariate normal distribution of the slope residuals (from Assumption 3). If there is no serial correlation in the u_{it}'s and they have a common variance σ_u^2, then the joint distribution simplifies to $f(y_1, ..., y_T|M_i, a; \sigma_u) = \Pi_{t=1}^{T}\{(1-D_{it})\log(1-\Phi(\frac{M_{it}\theta+Z_{it}a}{\sigma_u}))$ $+D_{it}\log\frac{1}{\sigma_u}\phi(\frac{y_{it}-(M_{it}\theta+Z_{it}a)}{\sigma_u})\}$. While the full-information MLE is more efficient under correct specification, this efficiency comes at the expense of less robustness (potential inconsistency when serial correlation is misspecified) and additional computation complexity.

Correlated-random-slopes Tobit Models

The assumption of random slopes may be too restrictive. In the same way that fixed effects models allow the intercept to be related to covariates (in contrast to random effects models), it may be preferable to allow the slope heterogeneity to itself be related to covariates. In this subsection, we consider a *correlated random-slopes Tobit model* in which slope heterogeneity is allowed to depend upon covariates, although not in a totally unrestricted way as would be the case for a fixed-effects-in-slopes model.[4] In particular, we adopt a familiar approach due to Chamberlain (1984) that has been widely utilized for non-linear models with correlated random effects (in intercepts).

Denote $X_{it} \equiv [X_{1i} \quad X_{2it} \quad X_{3t}]$, where X_1 contains time-invariant components, X_2 contains components that vary with both individuals and time, and X_3 contains individual-invariant components (such as year indicators or macroeconomic variables). Likewise, partition $Z_{it} \equiv [Z_{1i} \quad Z_{2it} \quad Z_{3t}]$. The time-invariant covariates are combined into $S_{1i} \equiv [X_{1i} \quad Z_{1i}]$, a $1 \times q_1$ vector, and time-and-individual-varying covariates into $S_{2it} \equiv [X_{2it} \quad Z_{2it}]$, a $1 \times q_2$ vector. The following assumption, allowing for correlation between the slopes and covariates, replaces Assumption 2:

Assumption 4. (Correlated random slopes — Chamberlain) *For each* $k \in \{1, \ldots, p_2\}$,

$$c_{ki} = S_{1i}\lambda_{1k} + \sum_{j=1}^{T} S_{2ij}\lambda_{2kj} + a_{ki}, \tag{2.6}$$

where the parameter vector λ_{1k} *is* $q_1 \times 1$ *and* λ_{2kj} *is* $q_2 \times 1$. c_{ki} *and* a_{ki} *are components of* $c_i \equiv (c_{1i}, \ldots, c_{p_2 i})$ *and* $a_i \equiv (a_{1i}, \ldots, a_{p_2 i})$*, respectively.*

In combination with the normality assumption on a_i (Assumption 3(ii)), Assumption 4 completely specifies the distribution of c_i conditional on the

[4]It is not obvious how one would estimate a fixed-effects-in-slopes model with censored data. Within-type transformations that are used for linear models are not applicable. Treating the c_i's as parameters to be estimated leads to an incidential-parameters problem in the censored-regression context. Moreover, the estimators of Honoré (1992) for the fixed-effects-in-intercept case do not immediately generalize.

covariates. Without loss of generality, assume that a constant term is included in either X_1 or Z_1 so that the a_{ki}'s have mean zero. Individual-invariant covariates, X_{3t} and Z_{3t}, are omitted from (2.6) since they would be collinear with the constant term.[5]

Substitution of (2.6) into (2.1) yields the latent-variable model

$$\begin{aligned} y_{it}^* &= X_{it}\beta + (Z_{it} \otimes S_{1i})\lambda_1 + \sum_{j=1}^{T}(Z_{it} \otimes S_{2ij})\lambda_{2j} + Z_{it}a_i + u_{it} \\ &= X_{it}\beta + W_{it}\lambda + Z_{it}a_i + u_{it} \end{aligned} \tag{2.7}$$

where $W_{it} \equiv (Z_{it} \otimes S_{1i} \;\; Z_{it} \otimes S_{2i1} \;\; Z_{it} \otimes S_{2i2} \;\; \cdots \;\; Z_{it} \otimes S_{2iT})$ is a $1 \times (q_1 p_2 + q_2 p_2 T)$ row vector. $\lambda_1 \equiv (\lambda_{11}' \;\; \lambda_{12}' \;\; \cdots \;\; \lambda_{1p_2}')'$ is a $q_1 p_2 \times 1$ column vector, $\lambda_{2j} \equiv (\lambda_{21j}' \;\; \lambda_{22j}' \;\; \cdots \;\; \lambda_{2p_2j}')'$ is a $q_2 p_2 \times 1$ column vector for each $j \in \{1, \ldots, T\}$, and $\lambda \equiv (\lambda_1' \;\; \lambda_{21}' \;\; \lambda_{22}' \;\; \ldots \;\; \lambda_{2T}')'$.

Re-defining $M_{it} \equiv [X_{it} \;\; W_{it}]$, the pooled Tobit estimator based upon (2.4) or the full-information likelihood estimator based upon (2.5) can be applied to the model in (2.7). The pooled estimator will be called the *correlated-random-slopes Tobit estimator*. The remarks regarding parameter identification made in Section 2.2.1 apply here as well, as do the relative tradeoffs between the pooled (partial-likelihood) estimator and the full-information estimator.

[5] In practice, any other covariates that would not be identified by a fixed-effects approach can be omitted from (2.6), as their associated parameters in (2.6) are not separately identified from the other parameters in the model. An example would be an element of S_2, such as age, that varies over time according to a fixed pattern and is not separately identified from time effects.

A potential problem with the Chamberlain-type relationship in Assumption 4 arises when T is relatively large, in which case the number of parameters to be estimated can become burdensome. A possible dimension-reduction solution is to make a further restriction within Assumption 4, specifically assuming that the slopes depend only upon the average of the time-varying covariates. This approach will be termed the Mundlak approach, in connection with the work by Mundlak (1978) on linear panel data models. The assumption for the Mundlak approach is given by

Assumption 5. (Correlated random slopes — Mundlak) *For each* $k \in \{1, \ldots, p_2\}$,

$$c_{ki} = \bar{S}_i \lambda_k + a_{ki}, \tag{2.8}$$

where $\bar{S}_i = [S_{1i} \;\; \frac{1}{T}\sum_{t=1}^{T} S_{2it}]$ *is a* $1 \times (q_1 + q_2)$ *row vector and* λ_k *is a* $(q_1 + q_2) \times 1$ *column vector.*

Substitution of (2.8) into (2.1) yields

$$\begin{aligned} y_{it}^* &= X_{it}\beta + (Z_{it} \otimes \bar{S}_i)\lambda + Z_{it}a_i + u_{it} \\ &= X_{it}\beta + W_{it}\lambda + \varepsilon_{it}, \end{aligned} \tag{2.9}$$

where $W_{it} = Z_{it} \otimes \bar{S}_i = (z_{1it}\bar{S}_i \quad z_{2it}\bar{S}_i \quad \cdots \quad z_{p_2 it}\bar{S}_i)$ is $1 \times (q_1 + q_2)p_2$ and $\lambda = (\lambda_1 \quad \lambda_2 \cdots \quad \lambda_{p_2})'$ $(q_1 + q_2)p_2 \times 1$. Under the Mundlak specification, the number of estimable coefficient parameters has been reduced to no more than $p_1 + q_1 p_2 + q_2 p_2$, as compared to $p_1 + q_1 p_2 + q_2 p_2 T$ parameters in the Chamberlain specification.

Partial Effect Estimators

Censored models can be used for two distinct data situations, (i) data-coding problems and (ii) corner solution problems. Data-coding arises when the outcome variable of interest is not documented for values below (bottom-coding) or above (top-coding) certain thresholds. In contrast, corner solution problems arise when a continuous outcome variable has positive probability at one or both ends of its support. As Wooldridge (2002) clarifies, the partial effects of interest for these two situations are quite different, with the empirical researcher usually concerned about effects on the latent (but possibly censored) dependent variable y^* in case (i) and effects on the observed dependent variable y (which includes the censored zero values) in case (ii).

Estimation of partial effects in the data-coding situation is straightforward since the latent variable is itself the object of interest. In the case of a continuous component $x_{t,j}$ of x_t, the partial effect of $x_{t,j}$ on the conditional expectation of y_t^* is given by the corresponding coefficient:

$$\frac{\partial E(y_t^*|x_t, c)}{\partial x_{t,j}} = \beta_j. \tag{2.10}$$

Other partial effects, including those for discrete variables and those for variables entering into interaction terms, can be evaluated in the same way as linear-regression models.

For corner-solution problems, estimation of partial effects is more difficult since the object of interest, y_t, has a conditional expectation that is non-linear in the covariates. For simplicity, consider the case in which the

Mundlak specification for the correlated random slopes is assumed (Assumption 5). The estimated conditional expectation, which uses the sample to integrate over the distribution of individual heterogeneity, is given by:

$$
\begin{aligned}
&\hat{E}(Y|X=x, Z=z) = \hat{E}_a(Y|X=x, Z=z, a) \qquad (2.11)\\
=&\frac{1}{N}\sum_{i=1}^{N}\Bigg[\left(x_t\hat{\beta} + (z_t \otimes \bar{S}_i)\hat{\lambda}\right)\Phi\left(\frac{x_t\hat{\beta} + (z_t \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z_t\hat{\Sigma}z_t')^{\frac{1}{2}}}\right)\\
&+\left(\hat{\sigma}_u^2 + z_t\hat{\Sigma}z_t'\right)^{\frac{1}{2}}\phi\left(\frac{x_t\hat{\beta} + (z_t \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z_t\hat{\Sigma}z_t')^{\frac{1}{2}}}\right)\Bigg].
\end{aligned}
$$

The partial effect of a continuous covariate $x_{t,j}$ is obtained by differentiating (2.11) with respect to $x_{t,j}$. For a discrete covariate, the partial effect can be estimated by evaluating (2.11) at different values of the discrete covariate and fixed values of the other covariates.[6]

Specification Tests

This subsection describes some specification tests that can be conducted within the context of the models described above. Since we are not introducing any new methodology with respect to testing, the ensuing discussion is relatively brief.

[6]The partial effects estimators will be asymptotically normal in cases when the underlying parameters are asymptotically normal, specifically when the slope heterogeneity does not have a degenerate distribution conditional on the factor variables. Either the delta method or standard clustered bootstrapping could be used to compute standard errors. If some of the individual-specific slopes have zero variance, the standard errors can be computed using the rescaled bootstrap procedure described in Section 2.1.1.

Testing random slopes versus correlated random slopes: The only difference between the random slopes model and the correlated random slopes model is the assumption made on the slope heterogeneity, which is unrelated to the covariates for the former and related to the covariates for the latter. The stricter random slopes assumption (Assumption 2) can be considered a special case of either the Chamberlain specification (Assumption 4) or the Mundlak specification (Assumption 5). For either case, one can test the random slope specification by estimating the correlated random slopes model and then directly testing whether the λ coefficients in equation (2.6) or (2.8) are jointly equal to zero.

Testing the randomness of slopes: In either the random slopes model or the correlated random slopes model, the heterogenous slope parameters are allowed to have a random component to them, denoted by a_i above. For the random slopes model, the null of homogenous slopes corresponds to the variance-covariance matrix Σ being equal to zero. Similarly, for the correlated random slopes model, the null of deterministic slopes (slopes being deterministic linear functions of covariates) also corresponds to the variance-covariance matrix Σ being equal to zero. Letting γ denote the stacked diagonal (variance) elements of Σ,[7] the null hypothesis of interest is $\gamma = 0$. Note that the variance parameters associated with the null hypothesis

[7] Recall that the variance corresponding to the constant term is not separately identified from σ_u^2.

lie on the boundary of the parameter space. As a result, the testing approach of Andrews (2001) should be applied.

2.2.2 Censored Least Absolute Deviation Estimation

This section relaxes the normality assumption used for the Tobit-type estimator of Section 2.2.1. Under a alternative conditional median assumption, a pooled version of the censored least-absolute deviations (CLAD) estimator of Powell (1984) can be used to estimate the model. Assumption 3 is replaced by

Assumption 6. (Zero conditional median)

$$Med(Z_{it}a_i + u_{it}|X_i, Z_i) = 0 \textit{ for } t = 1, 2, \ldots, T. \tag{2.12}$$

The normality assumption (Assumption 3) is a special case of the conditional median assumption, but Assumption 6 allows for non-normality and general forms of heteroskedasticity across both i and t.

Under Assumption 6, the CLAD-type estimator can be used to estimate the parameters of the three different models considered in Section 2.2.1: (i) random slopes, (ii) correlated random slopes with a Chamberlain specification, and (iii) correlated random slopes with a Mundlak specification.

Random slopes model (Assumption 2): Under the latent-variable model in equation (2.2), Assumption 6 implies

$$\text{Med}(y_{it}|X_i, Z_i) = \max\{0, X_{it}\beta + Z_{it}\lambda\}. \tag{2.13}$$

Chamberlain-type correlated random slopes model (Assumption 4): Under the latent-variable model in equation (2.7), Assumption 6 implies

$$\text{Med}(y_{it}|X_i, Z_i) = \max\{0, X_{it}\beta + (Z_{it} \otimes S_{1i})\lambda_1 + \sum_{j=1}^{T}(Z_{it} \otimes S_{2ij})\lambda_{2j}\}. \tag{2.14}$$

Mundlak-type correlated random slopes model (Assumption 5): Under the latent-variable model in equation (2.9), Assumption 6 implies

$$\text{Med}(y_{it}|X_i, Z_i) = \max\{0, X_{it}\beta + (Z_{it} \otimes \bar{S}_i)\lambda\}. \tag{2.15}$$

A common notation for the conditional median expressions in these three models can be obtained by defining M_{it} to be the full covariate vector in each of equations (2.13)–(2.15). Specifically, we have

$$M_{it} \equiv [X_{it} \quad Z_{it}] \tag{2.16}$$

for the random slopes model,

$$M_{it} \equiv [X_{it} \quad Z_{it} \otimes S_{1i} \quad Z_{it} \otimes S_{2i1} \quad Z_{it} \otimes S_{2i2} \quad \cdots \quad Z_{it} \otimes S_{2iT}] \tag{2.17}$$

for the Chamberlain-type correlated random slopes model, and

$$M_{it} \equiv [X_{it} \quad Z_{it} \otimes \bar{S}_i] \tag{2.18}$$

for the Mundlak-type correlated random slopes model. Note that estimation of the random slopes model here is just a pooled CLAD estimator of y_{it} on

$[X_{it} \;\; Z_{it}]$, in contrast to the MLE context where the pooled Tobit and random-slopes Tobit potentially yield different results due to the heteroskedasticity incorporated into the latter. Then, each of the specific models in (2.13)–(2.15) can be written as

$$\text{Med}(y_{it}|X_i, Z_i) = \max\{0, M_{it}\theta\}, \tag{2.19}$$

where θ is the stacked vector of parameters corresponding to the components of M_{it} defined in (2.16)–(2.18).

The pooled CLAD estimator $\hat{\theta}$, based upon the conditional median function (2.19), is obtained by minimizing the objective function[8]

$$S(\tilde{\theta}) \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| y_{it} - \max\left\{0, M_{it}\tilde{\theta}\right\} \right|$$

over $\tilde{\theta}$. The asymptotic properties of the pooled CLAD estimator follow directly from Powell (1984), with the exception that the asymptotic variance formula needs to incorporate clustering at the cross-sectional level:

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \rightarrow N\left(0, \frac{1}{4} A^{-1} B A^{-1}\right), \tag{2.20}$$

where

$$A = \sum_{t=1}^{T} E\left[f_t(0|M_{it}) 1\left(M_{it}\theta > 0\right) M_{it}' M_{it}\right] \tag{2.21}$$

and

$$B = E\left[\left(\sum_{t=1}^{T} \text{sign}(y_{it} - M_{it}\theta) 1(M_{it}\theta > 0) M_{it}'\right)\left(\sum_{t=1}^{T} \text{sign}(y_{it} - M_{it}\theta) 1(M_{it}\theta > 0) M_{it}\right)\right]. \tag{2.22}$$

[8] In the empirical section, we use the interior point optimization method of Koenker and Park (1996) and encounter no problems with local optima.

sign(x) is a indicator function equal to 1 if $x > 0$ and -1 if $x < 0$. $f_t(0|M_{it})$ is the conditional density of the error term $Z_{it}a_i + u_{it}$ at zero. As in the cross-sectional context, nonparametric estimation of conditional densities (required for estimation of A) can be avoided by implementing the bootstrap for statistical inference. For the panel version of the bootstrap, clustered re-sampling is used — drawing with replacement from $i \in \{1, \ldots, N\}$ and, for each draw, including all T periods of data in the bootstrap sample.

The pooled CLAD estimator is more robust than the pooled MLE estimator, as consistency of the former does not rely on parametetric or homoskedasticity assumptions. As such, a specification test for any of the parametric specifications of Section 2.2.1 can be based upon the difference between the pooled CLAD estimate vector and the pooled MLE estimate vector (coefficient parameters only, not variance parameters). Such a test is appropriate in both data-coding and corner-solution situations.

The pooled CLAD estimator is limited in terms of the partial effects that it can deliver. Without a distributional assumption on the heterogeneity a_i, it is impossible to estimate average partial effects of the type discussed in Section 2.2.1. For data-coding situations, partial effects on the outcome's conditional median can be estimated since $\mathrm{Med}(y^*_{it}|X_i, Z_i) = M_{it}\theta$.

2.3 Application: Household portfolio decisions

Household portfolio choice decisions are related to observed household heterogeneity (education, age, household income inflow, household asset composition, etc.) but also depend upon unobserved household heterogeneity (risk preference, knowledge in financial markets, future expenditure plans, etc.). While a traditional panel-data approach would assume that unobserved heterogeneity enters into the portfolio decision in an additively separable, we will use the slope heterogeneity models of this paper to allow unobserved heterogeneity to interact with observable household characteristics. Classical investment theory predicts, for instance, that the portfolio composition should depend upon an individual's age. In a model with no labor income and two assets (risky or safe), Samuelson (1969) and Merton (1969) show that the optimal share of safe assets for an investor with CRRA utility is a constant that depends only on individual's risk aversion rate, mean market equity premium, and the standard deviation of the risky asset return. When labor income is added to the model, as in Merton (1971), the optimal safe-asset ratio becomes a function of the present discounted value of future labor income PDV_t, total financial wealth FW_t, risk aversion rate γ, mean market equity premium μ, and the standard deviation of risky asset return σ_η:

$$\alpha_t = 1 - \frac{\mu}{\gamma\sigma_\eta^2}\left(1 + \frac{PDV_t}{FW_t}\right).$$

As a household's members get closer to retirement age, PDV_t decreases and FW_t increases from accumulated saving. Therefore, the optimal share of safe assets increases with age before retirement. The size of this effect is positively correlated with the risk aversion rate and the standard deviation of risky asset return and negatively correlated with the mean market equity premium. After retirement, the optimal share of safe assets is a fixed constant depending on these factors.

In this section, we consider data on household portfolio choice from the Netherlands. The dependent variable y_{it} is the share of safe assets in household i's portfolio at time t and is subject to two-sided censoring (left censoring at zero and right censoring at one).[9] Motivated by the stylized model described above, we specify a model that allows heterogeneity (both observed and unobserved) to interact with the observable age-profile variables. We focus on estimation of two specifications, the first a random-slopes model for the latent variable

$$y_{it}^{*} = [X_{1it} \quad X_{2t} \quad \text{AgeCategories}_{it} \quad 1]c_i + u_{it}, \tag{2.23}$$

and the second a correlated-random-slopes model for the latent variable

$$y_{it}^{*} = [X_{1it} \quad X_{2t}]\beta + [\text{AgeCategories}_{it} \quad 1]c_i + u_{it}. \tag{2.24}$$

The observed variable y_{it} is

$$y_{it} = \max\left\{\min\left\{Y_{it}^{*}, 1\right\}, 0\right\}. \tag{2.25}$$

[9]Models of intercept heterogeneity for this type of data have been studied in the literature, including the fixed-effects estimator for two-sided censored models of Alan, Honoré, and Leth-Petersen (2008) and the correlated random effects estimator for fractional response variabels of Papke and Wooldridge (2008). Extension of those methods to the case of heterogeneous slopes is left for future research.

The covariate vector X_1 contains household characteristic variables such as education, log financial assets, log total wealth, and log non-capital income, and X_2 contains systematic factors such as time trend and the AEX (Amsterdam Exchange index) mean premium rate and dispersion index in the 12 months preceding the survey start date.

The dataset is extracted from the DNB Household Survey of Netherlands, a longitudinal survey launched in 1993 that contains demographic and financial information for a large sample of Dutch individuals. Alessie, Hochguertel, and van Soest (2002) provide a detailed summary on the DNB dataset and Dutch household portfolio characteristics. For the estimation sample, we use households that appear in the DNB data at least three times in the period between 1993 and 2008.[10] We limit our households to those with heads older than 40 so as to limit the effects of risky non-financial assets, such as housing equity, on household financial portfolio position. We drop households that have missing data on household characteristics or financial behavior. The resulting sample has 11,954 total observations for 2,205 households. The dependent variable y_{it} is defined as the fraction of total financial wealth held in "clearly safe assets" by a household. "Clearly safe assets" are defined to include checking accounts, deposit books, saving or deposit accounts, and saving certificates, following Hochguertel (2003). Most of the censoring of y_{it} is due to right censoring at one (3,613 out of 11,954 (30.2%) fully invested in safe assets) rather than left censoring at zero (179 out of 11,954 (1.5%) fully

[10]Data from 2001 and 2002 are excluded due to obvious mis-coding issues.

Table 2.1: Descriptive statistics

	Mean	Std. Dev.
Proportion of safe assets (y_{it})	0.64	0.35
Age of household head	55.8	10.97
Education of household head:		
Secondary	0.11	0.31
Vocational	0.48	0.50
University	0.15	0.36
log(Financial assets)	10.17	1.68
log(Wealth)	11.78	1.56
log(Non-capital income)	9.47	3.80
Deposit interest rate	1.04	0.02
AEX annual return	1.12	0.24
AEX dispersion index	3.23	3.64

invested in risky assets). Table 1 provides summary statistics for the variables used in the data analysis.

For the empirical specifications in equations (2.23) and (2.24), the individual-varying covariates in X_{1it} are the education indicators (secondary, vocational, university), log(Financial assets), log(Wealth), and log(Non-capital income); the individual-invariant (but time-varying) covariates in X_{2t} are the AEX premium (defined as the AEX mean return minus the deposit rate), the AEX dispersion index (mean-to-variance ratio), and a linear time trend;[11]

[11] Daily closing prices for the AEX index were obtained from `http://finance.yahoo.com/q/hp?s=^AEX`. Deposit rates were obtained from `http://www.statistics.dnb.nl/index.cgi?lang=uk&todo=Rentes` (Table T1.3). AEX mean and variance calculations were based upon the daily closing values for the 12-month period preceding the start of the survey period. For example, the 1993 survey was conducted between November 1993 and April 1994, so the daily closing AEX index prices for the period between November 1992 and October 1993 were used.

and, there are six age categories within $AgeCategories_{it}$ (46-50, 51-55, 56-60, 61-65, 66-70, 71-75), with 41-45 as the omitted baseline category. Table 2 reports the coefficient estimates across several different model specifications and estimation approaches. Columns (1)–(4) correspond to MLE estimation of Tobit-type models, with (1) the pooled Tobit (homoskedasticity), (2) the random slopes Tobit, (3) the correlated random intercept Tobit, and (4) the correlated random slopes Tobit. Columns (5)–(7) correspond to CLAD estimation, with (5) the pooled CLAD (equivalent to the random slopes CLAD), (6) the correlated random intercept CLAD, and (7) the correlated random slopes CLAD. For the correlated random slopes models (columns (4) and (7)), the Mundlak specification of Assumption 5 is used for the six age-category variables. The Mundlak coefficients associated with the age categories, which will be discussed later, are shown below in Table 5. The random slopes Tobit and correlated random slopes Tobit (columns (2) and (4), respectively) were estimated under the assumption that Σ is diagonal.[12] To allow for serial correlation in the u_{it} disturbances, partial maximum likelihood estimators are used throughout.

The pooled Tobit (column (1)) and random-slopes Tobit (column (2)) estimates of Table 1 are very similar to each other. The secondary and university education indicators are significantly negative, indicating that highly educated households are more likely to hold risky assets. This finding is con-

[12]The estimates allowing for non-zero off-diagonal elements are qualitatively very similar to those reported here and are available from the authors.

sistent with these individuals having both higher discounted future income and also more knowledge about risky financial instruments. Financial assets have a significantly negative impact, indicating decreasing relative risk aversion. Total wealth and non-capital income are negative and significant, but the statistical significance of these variables disappears when correlated heterogeneity is allowed in either the intercept (column (3)) or the age slopes (column (4)). The AEX equity premium has a negative impact on safe-asset holding, as expected. The AEX variability has an unexpected negative sign, which may be arising due to the normalization that we are using (division by the AEX mean return). For the age categories, the coefficient estimates of the first two age groups are insignificant, indicating similar portfolio composition for household heads aged 40-55, whereas the estimates for the older age categories are positive, significant, and increasing in magnitude at higher ages. The correlated random intercept model in column (3) yields very similar results for most variables, with a few exceptions: the magnitude of the financial assets coefficient declines and the significant of wealth and non-capital income disappear. These differences are also seen in the correlated random slopes model of column (4). Finally, we note that the CLAD estimates of columns (5)–(7) have similar sign and significance to their corresponding Tobit estimates in the table. As expected, the CLAD standard errors are larger than those obtained by MLE.

Table 3 reports the standard deviation estimates associated with the random slopes Tobit specification (column (2) of Table 2). For comparison

Table 2.2: Portfolio-composition regression results

	Tobit				CLAD		
	(1) Pooled	(2) Random Slopes	(3) Correlated Random Intercept	(4) Correlated Random Slopes	(5) Pooled	(6) Correlated Random Intercept	(7) Correlated Random Slopes
Secondary education	-0.0757**	-0.0653**	-0.0703**	0.0833	-0.0730*	-0.0700*	0.0482
	(0.0296)	(0.0295)	(0.0298)	(0.0569)	(0.0383)	(0.0387)	(0.0686)
Vocational education	-0.0298	-0.0192	-0.0256	0.0813**	-0.0058	-0.0113	0.0698
	(0.0199)	(0.0200)	(0.0198)	(0.0382)	(0.0252)	(0.0257)	(0.0458)
University education	-0.0669***	-0.0556**	-0.0571**	0.0671	-0.0356	-0.0346	0.0616
	(0.0249)	(0.0248)	(0.0253)	(0.0471)	(0.0343)	(0.0356)	(0.0633)
log(FA)	-0.1449***	-0.1429***	-0.1228***	-0.1170***	-0.1846***	-0.1663***	-0.1695***
	(0.0077)	(0.0079)	(0.0106)	(0.0101)	(0.0105)	(0.0163)	(0.0148)
log(Wealth)	-0.0245***	-0.0258***	-0.0060	-0.0082	-0.0158	0.0031	0.0095
	(0.0083)	(0.0083)	(0.0119)	(0.0114)	(0.0117)	(0.0140)	(0.0126)
log(NCI)	-0.0034**	-0.0036**	-0.0009	-0.0011	-0.0032	0.0000	-0.0002
	(0.0016)	(0.0017)	(0.0012)	(0.0012)	(0.0023)	(0.0018)	(0.0019)
Age 46-50	-0.0012	0.0040	-0.0018		-0.0020	-0.0023	
	(0.0178)	(0.0176)	(0.0180)		(0.0265)	(0.0268)	
Age 51-55	0.0244	0.0253	0.0267		0.0301	0.0304	
	(0.0198)	(0.0194)	(0.0201)		(0.0281)	(0.0299)	
Age 56-60	0.0519**	0.0570**	0.0524**		0.0499	0.0448	
	(0.0219)	(0.0219)	(0.0225)		(0.0313)	(0.0325)	
Age 61-65	0.1293***	0.1344***	0.1306***		0.1181***	0.1239***	
	(0.0231)	(0.0235)	(0.0239)		(0.0297)	(0.0303)	
Age 66-70	0.2503***	0.2693***	0.2519***		0.2358***	0.2341***	
	(0.0257)	(0.0275)	(0.0260)		(0.0312)	(0.0421)	
Age 71-75	0.2854***	0.3094***	0.2922***		0.2977***	0.2936***	
	(0.0307)	(0.0335)	(0.0315)		(0.0418)	(0.0421)	
AEX prem.	-0.0641***	-0.0644***	-0.0626***	-0.0553***	-0.0869***	-0.0781***	-0.0730***
	(0.0167)	(0.0170)	(0.0167)	(0.0169)	(0.0256)	(0.0264)	(0.0260)
AEX disp. index	-0.0024**	-0.0027**	-0.0038***	-0.0035***	-0.0031*	-0.0040**	-0.0037**
	(0.0011)	(0.0011)	(0.0012)	(0.0012)	(0.0017)	(0.0017)	(0.0017)
Time trend (year)	-0.0134***	-0.0134***	-0.0136***	-0.0151***	-0.0115***	-0.0116***	-0.0142***
	(0.0024)	(0.0015)	(0.0015)	(0.0015)	(0.0020)	(0.0021)	(0.0023)

Notes: The dependent variable is defined as the share of safe assets (between 0 and 1) in the household portfolio. Tobit standard errors are computed with the MLE asymptotic cluster-adjusted formula. CLAD standard errors are computed with a clustered bootstrap (5000 replications). Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

purposes, the pooled Tobit estimate of the residual standard deviation σ_u (0.4120) is shown. The corresponding estimate for the random slopes specification is 0.3717, meaning that roughly 18.6% of the overall residual variance comes through the random slopes. Since we find several standard deviation estimates close to zero, the re-scaled bootstrap procedure of Andrews (1999, 2001) is used; the reported standard error estimates are based upon subsample sizes of $0.25N$.[13] There is strong evidence that the random slopes Tobit model is preferable to the pooled (homoskedasticity) specification, with significant standard-deviation estimates associated with two of the educational categories (secondary, vocational), five of the age categories, and the time trend.

Table 4 reports the sample average partial effect estimates for the full sample and the uncensored sample for selected explanatory variables.

Since the portfolio decision problem can have corner solutions (at $y_{it} = 0$ for all risky assets and $y_{it} = 1$ for all safe assets), the coefficient estimates of Table 2 do not directly provide the partial effects of interest. Instead, we use the following two-sided censoring generalization of equation (2.11), where a denotes the left-censoring point ($a = 0$ here) and b denotes the right-censoring

[13] We also calculated re-scaled bootstrap standard errors for the coefficients reported in Table 2, but we found essentially no differences between the re-scaled and unscaled bootstrap results.

Table 2.3: Standard-deviation estimates for the pooled and random slopes Tobit models

	(1) Pooled	(2) Random Slopes		(1) Pooled	(2) Random Slopes
σ_u	0.4120***	0.3717***			
	(0.0056)	(0.0102)			
Secondary education		0.0970**	Vocational education		0.1258***
		(0.0477)			(0.0357)
University education		0.0000#	log(Financial assets)		0.0000#
log(Wealth)		0.0000#	log(Non-capital income)		0.0000#
Age 46-50		0.1352***	Age 51-55		0.0000#
		(0.0377)			
Age 56-60		0.1488***	Age 61-65		0.1395***
		(0.0418)			(0.0412)
Age 66-70		0.2265***	Age 71-75		0.2353***
		(0.0441)			(0.0559)
AEX premium		0.0000#	AEX dispersion index		0.0000#
Time trend (year)		0.0079***			
		(0.0027)			

Notes: Each estimate corresponds to the standard deviation associated with the random slope of the particular variable (square root of the Σ diagonal element). Standard errors are calculated using the re-scaled bootstrap of Andrews (1999, 2001) based upon 25% of the original sample size (5000 replications). # indicates a standard deviation estimate that was equal to zero to at least four decimal places. Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

point ($b = 1$ here):

$$\begin{aligned}
&\hat{E}(Y|X = x, Z = z) \\
&= \frac{1}{N}\sum_{i=1}^{N}\Bigg\{ \left(x\hat{\beta} + (z \otimes \bar{S}_i)\hat{\lambda}\right)\left[\Phi\left(\frac{b - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right) - \Phi\left(\frac{a - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right)\right] \\
&- \left(\hat{\sigma}_u^2 + z\hat{\Sigma}z'\right)^{\frac{1}{2}}\left[\phi\left(\frac{b - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right) - \phi\left(\frac{a - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right)\right] \\
&+ a\Phi\left(\frac{a - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right) + b\left(1 - \Phi\left(\frac{b - x\hat{\beta} - (z \otimes \bar{S}_i)\hat{\lambda}}{(\hat{\sigma}_u^2 + z\hat{\Sigma}z')^{\frac{1}{2}}}\right)\right)\Bigg\}.
\end{aligned} \tag{2.26}$$

Note that equation (2.11) is a special case of equation (2.26) with $a = 0$ and $b \to \infty$.

The partial effect of a continuous covariate, evaluated at given x and z, is obtained by differentiating (2.26) with respect to the covariate. To provide an average partial effect estimate, one can take this covariate-value-specific partial effect for each observation and then take an average. These average partial effects are reported in Table 4. The table reports average partial effects corresponding to each of the four Tobit specifications, columns (1)–(4) from Table 2. The last column reports the difference in partial effects between the correlated random intercept model and the correlated random slopes model. Overall, the signs and statistical significance are quite similar to the coefficient results reported in Table 2. As for the coefficient estimates in Table 2, we note that the partial effects of wealth and non-capital income become insignificant once heterogeneity is allowed to be correlated with observables. The importance of financial assets also declines somewhat for the correlated random intercept and correlated random slopes models. The last column of the table highlights some important differences between the correlated random intercept estimates and the correlated random slopes estimates. Specifically, the partial effects for the three oldest age categories (61-65, 66-70, 71-75) are significantly larger for the correlated random slopes model. Allowing for slope heterogeneity to be correlated with observables yields partial effects that are roughly 1 percentage point higher for the 61-65 age group and 4 percentage points higher for the 66-70 and 71-75 age groups. This model, therefore, predicts a steeper age profile with respect to safe-asset holding.

Finally, we report the coefficients associated with the Mundlak specification (2.8) for the correlated random slopes Tobit (specification (4) from Table 2) and the correlated random slopes CLAD (specification (7) from Table 2). Recall that the model in (2.24) allowed the slope heterogeneity on the age-category variables to be correlated with other observables in the model. For each age category, then, the table reports the coefficient estimates associated with these observables in the Mundlak specification and, for the Tobit model, the estimated standard deviation of the residual. As expected, these estimated residual standard deviations are similar in magnitude but somewhat smaller than the corresponding estimates for the random slopes model in Table 3. The most interesting pattern that emerges from Table 5 is that the correlation of the slope heterogeneity with observables becomes much stronger and more statistically significant for later ages. For instance, the association with university education and log(Financial assets) is nearly twice as large for the Tobit slope in the age 66-70 category as it is in the age 61-65 category. Overall, the strong significance of many of the covariates within Table 5 pro-

Table 2.4: Average partial effect estimates for Tobit models

	(1) Pooled	(2) Random Slopes	(3) Correlated Random Intercept	(4) Correlated Random Slopes	Difference between (3) and (4)
log(Financial assets)	-0.0926***	-0.0914***	-0.0802***	-0.0760***	-0.0043**
	(0.0146)	(0.0145)	(0.0068)	(0.0064)	(0.0019)
log(Wealth)	-0.0157***	-0.0165***	-0.0039	-0.0053	0.0014
	(0.0055)	(0.0056)	(0.0078)	(0.0074)	(0.0014)
log(Non-capital income)	-0.0022**	-0.0023**	-0.0006	-0.0007	0.0001
	(0.0010)	(0.0011)	(0.0008)	(0.0008)	(0.0002)
Age 46-50	-0.0008	0.0015	-0.0012	0.0083	-0.0095***
	(0.0108)	(0.0107)	(0.0117)	(0.0112)	(0.0039)
Age 51-55	0.0156	0.0162	0.0175	0.0206	-0.0032
	(0.0124)	(0.0122)	(0.0131)	(0.0127)	(0.0035)
Age 56-60	0.0332**	0.0354***	0.0342***	0.0370***	-0.0028
	(0.0142)	(0.0144)	(0.0147)	(0.0144)	(0.0037)
Age 61-65	0.0826***	0.0848***	0.0853***	0.0972***	-0.0119***
	(0.0186)	(0.0190)	(0.0154)	(0.0159)	(0.0045)
Age 66-70	0.1599***	0.1689***	0.1646***	0.2050***	-0.0405***
	(0.0284)	(0.0301)	(0.0165)	(0.0184)	(0.0077)
Age 71-75	0.1824***	0.1955***	0.1909***	0.2308***	-0.0399***
	(0.0333)	(0.0361)	(0.0200)	(0.0244)	(0.0111)
AEX premium	-0.0410***	-0.0412***	-0.0409***	-0.0359**	-0.0050*
	(0.0118)	(0.0120)	(0.0109)	(0.0110)	(0.0026)
AEX dispersion index	-0.0016**	-0.0017**	-0.0025***	-0.0023***	-0.0002
	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0002)
Time trend (year)	-0.0079***	-0.0080***	-0.0082***	-0.0089***	0.0007***
	(0.0014)	(0.0014)	(0.0008)	(0.0008)	(0.0002)

Notes: Average partial effects are sample averages of the estimated partial effect at each observation. Standard errors are computed with a clustered bootstrap (5000 iterations). Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

vides evidence against the correlated random intercept specification; a formal Wald test based on either the Tobit or CLAD estimates results in overwhelming rejection. The correlated random slopes specification captures important interactions between observables and the effects (slopes) of age on portfolio composition choice. This importance is also reflected in the significantly larger partial effects for older households that had been found in Table 4, as compared to more restrictive specifications.

Table 2.5: Mundlak coefficient estimates for the correlated random slopes models

	(4) Tobit Age 46-50	(7) CLAD Age 46-50	(4) Tobit Age 51-55	(7) CLAD Age 51-55	(4) Tobit Age 56-60	(7) CLAD Age 56-60
Secondary education	-0.0739	-0.0414	-0.0890	-0.0650	-0.2043**	-0.2071
	(0.0672)	(0.0880)	(0.0729)	(0.0935)	(0.0846)	(0.1069)
Vocational education	-0.0653	-0.0708	-0.0882*	-0.0477	-0.1294**	-0.0736
	(0.0470)	(0.0622)	(0.0494)	(0.0651)	(0.0558)	(0.0762)
University education	-0.0752	-0.1122	-0.0199	0.0116	-0.1301*	-0.0419
	(0.0568)	(0.0774)	(0.0604)	(0.0781)	(0.0692)	(0.1002)
log(Financial assets)	0.0389**	0.0594**	0.0111	0.0390	-0.0066	-0.0373
	(0.0195)	(0.0278)	(0.205)	(0.0296)	(0.0245)	(0.0372)
log(Wealth)	-0.0146	-0.0337	-0.0136	-0.0405	0.0056	0.0016
	(0.0199)	(0.0287)	(0.0211)	(0.0328)	(0.20245)	(0.0385)
log(Non-capital income)	-0.0149*	-0.0238*	-0.0088	-0.0111	-0.0102	-0.0150
	(0.0080)	(0.0133)	(0.0075)	(0.0110)	(0.0091)	(0.0127)
σ_a	0.1164***		0.0000		0.1653***	
	(0.0366)		(0.0001)		(0.0339)	

	(4) Tobit Age 61-65	(7) CLAD Age 61-65	(4) Tobit Age 66-70	(7) CLAD Age 66-70	(4) Tobit Age 71-75	(7) CLAD Age 71-75
Secondary education	-0.2423***	-0.2060*	-0.3297***	-0.3176**	-0.2440**	0.0417
	(0.0906)	(0.1123)	(0.0998)	(0.1453)	(0.1127)	(0.2480)
Vocational education	-0.1266**	-0.0975	-0.1765**	-0.1726**	-0.2846***	-0.1800
	(0.0572)	(0.0689)	(0.0631)	(0.0865)	(0.0817)	(0.1401)
University education	-0.1755**	-0.1196	-0.3302***	-0.2065**	-0.4384***	-0.4684***
	(0.0783)	(0.1025)	(0.0822)	(0.1152)	(0.1041)	(0.1826)
log(Financial assets)	-0.0794***	-0.0578	-0.1242***	-0.1288***	-0.1093	-0.1117
	(0.0242)	(0.0375)	(0.0270)	(0.0466)	(0.0335)	(0.0709)
log(Wealth)	0.0362	-0.0345	0.0489	-0.0121	0.0286	-0.0.0234
	(0.0267)	(0.0435)	(0.0325)	(0.0616)	(0.0401)	(0.0977)
log(Non-capital income)	-0.0059	-0.0128	0.0143	0.0072	0.0088	0.0103
	(0.0095)	(0.0122)	(0.0104)	(0.0173)	(0.0128)	(0.0269)
σ_a	0.1716***		0.2416***		0.2543***	
	(0.0338)		(0.0354)		(0.0452)	

Notes: σ_a denotes the standard deviation random component associated with a given age category (square root of the Σ diagonal element). Tobit standard errors are computed with the MLE asymptotic cluster-adjusted formula. CLAD standard errors are computed with a clustered bootstrap (5000 replications). Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

Chapter 3

Kernel-based Specification Tests for Models with Endogenous Covariates

3.1 Introduction

The estimation of econometric models with endogenous covariates has long been of interest to econometricians and applied researchers. The objects of interest are typically structural functions and average/quantile partial effects rather than the conditional expectation of the outcome variable as in a model with exogenous covariates. In classic parametric models, structural functions and partial effects are identified through the estimation of unkown parameters, however the identification in such models is dependent on the particular parametric specification. In recent years, researchers have developed nonparametric and semiparametric estimation procedures for estimation of structural models with endogenous covariates, including Newey and Powell (2003), Ai and Chen (2003), Hall and Horowitz (2005), Newey et al. (1999), Blundell and Powell (2003), Imbens and Newey (2009), among others. The first three papers listed above use the classic zero conditional mean assumption of the error term given instruments to estimate the structural function while the other three uses the "control function/variable" approach. The assumptions required for the two identification strategies are not nested within each other. See Blundell and Powell (2003) for detailed discussion on the two identification strategies.

This paper focuses on testing the specification of structural models with endogenous covariates, given the existence of valid instruments. Donald, Imbens and Newey (2003), Kitamura et al. (2004) and Horowitz (2006) propose tests for parametric structural functions against nonparametric alternatives. Their tests are based on the conditional mean assumption of the error term given instruments. In this paper, I utilize the idea of control function/variable approach to develop nonparametric and semiparametric specification tests for structural models with endogenous covariates. The tests are developed for the validity of nonparametric specifications with a restricted set of regressors and of the semiparametric single index specification. I also extend the methodology to provide an alternative for testing parametric specifications employing the control function approach.

All test statistics to be discussed are based on kernel methods that date back to Rosenblatt (1956), Nadaraya (1964) and Waston (1964). Such methods have been used and extended in numerous ways in economic applications. One such extension is to test the specification of restrictive parametric or semiparametric models against the alternative of less restrictive nonparametric models. Some recent examples include, Yatchew (1992), Hardle and Mammen (1993), Whang and Andrews (1993), Fan and Lee (1996), Ait-Sahalia et al. (2001) and Stute and Zhu (2005). These papers typically assume that covariates are exogenous and are often based on the distance between the parametric and semiparametric or nonparametric alternative model. This paper extends this literature to allow for endogenous covariates in the model. Since in such cases the models are structural and do not have a conditional expectation interpretation a new apporach is needed to compare the models. In our case we show that we can test the hypotheses of interest by testing a conditional moment restriction involving a control function that accounts for endogeneity of certain covariates.

The remainder of the paper is organized as follows. In Section 3.2 we discuss the benchmark test for nonparametric restrictions, or whether or not a subset of covariates belong in a structural model. We also illustrate why existing tests based on conditional expectation comparisons are not useful in the endogenous variable case and then provide the test statistic and asymptotic theory. In Section 3.3 we extend the tests for parametric and semiparametric single index model specifications. All test statistics are constructed based on the control variable approach as in Blundell and Powell (2003) and the "leave-one-out" type kernel U-statistic as in Fan and Lee (1996). In Section 3.4 we examine the small sample properties of the proposed tests with simulated data. And in Section 3.5 we conclude the paper and discusses directions of future research.

3.2 Benchmark Test for Nonparametric Restrictions

3.2.1 Motivation

To motivate the tests that deal with endogenous covariates we first discuss the problem of specification testing with exogenous covariates and then show how the usual approach to testing needs to be adjusted when endoegenous covariates are present. Consider the following model,

$$Y = g(X) + \varepsilon,$$

$$E(\varepsilon|X) = 0,$$

where Y is the scalar outcome variable, $X \in R^d$ the covariate and ε the error term. The function $g(.) : R^d \to R$ is unknown. The covariate X is assumed to be exogenous for now and by the zero conditional mean assumption, $g(X)$ is equal to the conditional expectation of the outcome variable $E(Y|X)$.

Let the covariate X be partitioned with subsets X_1 and X_2, $X = (X_1\ X_2)$. In this section, we focus on testing whether the subset of covariate X_2 belongs to the function $g(.)$, or more specifically, whether X_2 influences the outcome Y after controlling for the subset of covariates X_1. The subset X_1 is referred to as the restricted set of covariates. Let d_r be the dimension of X_1 and $G(.)$ an unknown function mapping from R^{d_r} to the real line. The null hypothesis we are interested in testing is then

$$H_0^a : g(X) = G(X_1) \text{ a.e. for some } G(.) : R^{d_r} \to R, \text{ against the alternative}$$
$$H_1^a : g(X) \neq G(X_1) \text{ a.e. for all } G(.) : R^{d_r} \to R.$$

Under the null hypothesis, $Y = G(X_1) + u$. And by the exogeneity assumption on X, it follows that $G(X_1) = E(Y|X_1)$. Therefore, the null hypothesis H_0^a and the alternative hypothesis H_1^a are equivalent to

$$H_0'^a : E(Y|X) = E(Y|X_1) \text{ a.e. and}$$
$$H_1'^a : E(Y|X) \neq E(Y|X_1) \text{ a.e.}$$

in the model with exogenous covariates. $H_0'^a$ and $H_1'^a$ actually examine the goodness-of-fit of the restricted conditional expectation function $E(Y|X_1)$. Specification tests in Fan and Lee (1996) and Ait-Sahalia et al. (2001) for example are based on hypotheses $H_0'^a$ and $H_1'^a$.

Things are different when one or more of the covariates is endogenous. In such cases tests examining the goodness-of-fit of the restricted conditional expectations no longer provides a useful way of examining the specifications of the model since the model is no longer defined in terms of conditional expectations and instead has a structural interpretation. To see this suppose $X = (X_1, X_2, Y_2)$, where X_1, X_2 are exogenous covariates with $E(\varepsilon|X_1, X_2) = 0$ and Y_2 an endogenous covariate with $E(\varepsilon|Y_2) \neq 0$ and as above suppose that one is interested in testing whether X_2 can be eliminated from the model. Two illustrative examples below show why the usual testing approach based on conditional mean comparisons is not suitable for examining this hypothesis in this situation due to the presence of the endoegenous covariate Y_2.

Example 1: H_0^a is satisfied but $H_0'^a$ is not

The outcome variable $Y = X_1 + Y_2 + \varepsilon$, where the endogenous variable satisfies $Y_2 = X_1 + X_2 + Z_3 + V$. X_1, X_2, Z_3 are all scalar exogenous random variables independent of each other. Z_3 takes -1 and 1 with probability one-half (or any distribution symmetric about zero). Suppose $E(V|X_1, X_2, Z_3) = 0$, $\varepsilon = \frac{1}{2}V + e$, $E(e|V) = 0$. Then $E(\varepsilon|X_1, X_2, Y_2) = \frac{1}{2}(Y_2 - X_1 - X_2)$. We see that the structural model does not depend on X_2 but in this case the conditional mean does depend on X_2as $E(Y|X_1, X_2, Y_2) = \frac{1}{2}X_1 - \frac{1}{2}X_2 + \frac{3}{2}Y_2$.

Example 2: $H_0'^a$ is satisfied but H_0^a is not

The outcome variable $Y = X_1 + \frac{1}{2}X_2 + Y_2 + \varepsilon$ where the endogenous variable satisfies $Y_2 = X_1 + X_2 + Z_3 + V$. X_1, X_2, Z_3 are distributed the same as in Example 1. Again suppose $E(V|X_1, X_2, Z_3) = 0$, $\varepsilon = \frac{1}{2}V + e$, $E(e|V) = 0$. Then $E(\varepsilon|X_1, X_2, Y_2) = \frac{1}{2}(Y_2 - X_1 - X_2)$. We see that H_0^a is not satisfied but $H_0'^a$ is as $E(Y|X_1, X_2, Y_2) = \frac{1}{2}X_1 + \frac{3}{2}Y_2$.

With a similar argument we could show that existing specification tests for the semiparametric single index specification based on comparison of conditional expectations could not be used in situations where endogenous covariates present as well. We will discuss in the next section how single index specification tests could be performed in models with endogenous covariates.

3.2.2 Test Statistic and Asympotic Property

To develop tests that are useful in situations with endogenous covariates consider the structural model,

$$Y = g(X) + \varepsilon = g(X_1, X_2, Y_2) + \varepsilon, \tag{3.1}$$

$$Y_2 = H(Z) + V = H(X_1, X_2, Z_3) + V. \tag{3.2}$$

The outcome Y is a scalar variable. The unknown structural function $g(.) : R^{p_1+p_2+q} \rightarrow R$ for the outcome variable is a function of p_1 and p_2 dimensional exogenous covariates X_1 and X_2 and q dimensional endogenous covariate Y_2. ε is the error term or unobserved determinant of Y. The endogenous covariate Y_2 is continuous. It is further assumed to be additive of an unknown function $H(.) : R^{p_1+p_2+p_3} \rightarrow R$ and an unobserved error term V. $H(.)$ is a function of instrument, including X_1, X_2 and a p_3 dimensional extra exogenous covariate Z_3. The error terms ε and V are correlated with each other; ε is single dimensional and V is q dimensional. Given this setup, the hypothesis of interest is whether the subset of covariates X_2 can be omitted from the structural function g. The null and alternative hypotheses are,

$$H_0^a : g(X_1, X_2, Y_2) = G(X_1, Y_2) \text{ a.e. for some } G(.) : R^{p_1+p_2} \rightarrow R,$$
$$H_1^a : g(X_1, X_2, Y_2) \neq G(X_1, Y_2) \text{ a.e. for all } G(.) : R^{p_1+p_2} \rightarrow R.$$

Newey, Powell, and Vella (1999), Blundell and Powell (2003, 2004), Imbens and Newey (2009) and others study nonparametric estimation methods for models with endogenous covariates using the control function approach. Following Blundell and Powell (2003), assume here that

$$E(V|Z) = 0, \tag{3.3}$$
$$E[\varepsilon|Z, V] = E[\varepsilon|V]. \tag{3.4}$$

Then in the unrestricted model in (3.1)-(3.2),

$$
\begin{aligned}
E[Y|X,V] &= E[E[Y|X,Z_3,V]|X,V] = g(X) + E[E[\varepsilon|X,Z_3,V]|X,V] \\
&= E[g(X) + E[\varepsilon|Z,V]|X,V] = E[g(X) + E[\varepsilon|V]|X,V] \qquad (3.5) \\
&= g(X) + E[\varepsilon|V] = g(X) + h(V).
\end{aligned}
$$

The third equality holds because Y_2 is a function of Z and V, while the fourth holds because of the assumption in (3.4). Therefore the structural function satisfies

$$g(X) = \int (E(Y|X,V) - E[\varepsilon|V])\, \mathrm{d}F_V(V) = \int E(Y|X,V)\mathrm{d}F_V(V), \qquad (3.6)$$

where $F_V(V)$ is the cumulative distribution function of V. The second equality holds by the law of iterated expectation and the zero mean assumption on ε. The error term V is called the control variable or control function. By equation (3.6), Blundell and Powell (2003) propose to estimate the structural function $g(X)$ by the nonparametric sample analog of $\int E(Y|X,V)\mathrm{d}F_V(V)$, i.e. integrating the estimated conditional expectation of Y conditional on X and $\hat{V}$ over the empirical distribution of $\hat{V}$, where $\hat{V}$ is the first step residual estimator from the nonparametric estimation of equation (3.2).

The conditional mean assumptions in (3.3) and (3.4) are imposed for the structural function identification using the control function approach. Identification of the structural function is also possible under the assumption that

$$E(\varepsilon|Z) = 0, \qquad (3.7)$$

as proposed by Newey and Powell (2003) and Hall and Horowitz (2005) among others. The assumption in equation (3.7) and those in (3.3) and (3.4) are not nested within each other. Donald, Imbens and Newey (2003), Tripathi and Kitamura (2003) and Horowitz (2006) propose tests that study whether $g(.)$ is of certain parametric functional form based on the assumption in (3.7). I will discuss an alternative parametric test using the control function approach in the next section.

Parallel to (3.5), one can also note that under the null hypothesis,

$$\begin{aligned} E[Y|X_1, Y_2, V] &= E[g(X)|X_1, Y_2, V] + h(V), \\ &= G(X_1, Y_2) + h(V). \end{aligned}$$

Let $\xi = Y - E(Y|X_1, Y_2, V)$. It is obvious that under the null hypothesis $\xi = \varepsilon - E[\varepsilon|V]$ and hence $E[\xi|X] = E[E[\xi|X, V]|X] = 0$. Under the alternative $E[\xi|X] = g(X) - E[g(X)|X_1, Y_2] \neq 0$. Define $W = (X_1, Y_2, V)$. Let $a(W)$ and $a(X)$ be positive weighting functions depending on W and X respectively. The expression $E[\xi a(W)E[\xi a(W)|X]a(X)] = E[E[\xi a(W)|X]^2 a(X)] = 0$ when the null is true and it is positive otherwise. The test statistic to be used is an estimator of $E[\xi f(W)E[\xi f(W)|X]]$ with the particular weighting functions $f(W)$ and $f(X)$ which are probability density functions of W and X respectively.

Assumption 3.2.1. *Let $\{Y_i, X_{1i}, X_{2i}, Y_{2i}, Z_{3i}\}_{i=1}^n$ be independent and identically distributed sample observations on $\{Y, X_1, X_2, Y_2, Z_3\} \in R \times R^{p_1} \times R^{p_2} \times R^q \times R^{p_3}$. Denote $X_i = (X_{1i}, X_{2i}, Y_{2i})$ and $Z_i = (X_{1i}, X_{2i}, Z_{3i})$, $i = 1, 2, ...n$.*

Let $d_0 = p_1 + p_2 + p_3$ be the dimension of Z, $K_0(.) : R^{d_0} \to R$ the kernel function and h_0 the bandwidth. First we estimate the unknown function $H(.)$

nonparametrically by the leave-one-out kernel estimator

$$\hat{H}(Z_i) = \frac{\sum_{j\neq i}^{n} K_0\left(\frac{Z_i - Z_j}{h_0}\right) Y_{2j}}{\sum_{j\neq i}^{n} K_0\left(\frac{Z_i - Z_j}{h_0}\right)}$$

The estimator is $\sqrt{nh_0^{d_0}}$ consistent. Define $V_i = Y_{2i} - H(Z_i)$ and $\hat{V}_i = Y_i - \hat{H}(Z_i)$. $\hat{V}_i$ is the first step nonparametric estimator of the error term V evaluated at Z values for observation i.

Next we define estimators for the density function $f(X_1, Y_2, V)$ and the conditional expectation function $M(X_1, Y_2, V) = E(Y|X_1, Y_2, V)$. Let $W_i = (X_{1i}, Y_{2i}, V_i)$, $d_1 = p_1 + 2q$. Let $K_1(.) : R^{d_1} \to R$ be the kernel function and h_1 the bandwidth. Estimators

$$\begin{aligned}
\hat{f}(W_i) &= \frac{1}{(n-1)h_1^{d_1}} \sum_{j\neq i} K_1\left(\frac{W_i - W_j}{h_1}\right) \\
&= \frac{1}{(n-1)h_1^{d_1}} \sum_{j\neq i} K_1\left(\frac{X_{1i} - X_{1j}}{h_1}, \frac{Y_{2i} - Y_{2j}}{h_1}, \frac{V_i - V_j}{h_1}\right), \quad \text{and} \\
\hat{M}(W_i) &= \frac{1}{\hat{f}\,(W_i)} \frac{1}{(n-1)h_1^{d_1}} \sum_{j\neq i} K_1\left(\frac{W_i - W_j}{h_1}\right) Y_j \\
&= \frac{1}{\hat{f}\,(X_{1i,} Y_{2i}, V_i)} \frac{1}{(n-1)h_1^{d_1}} \sum_{j\neq i} K_1\left(\frac{X_{1i} - X_{1j}}{h_1}, \frac{Y_{2i} - Y_{2j}}{h_1}, \frac{V_i - V_j}{h_1}\right) Y_j.
\end{aligned}$$

are $\sqrt{nh_1^{d_1}}$ consistent estimators of $f(W_i)$ and $M(W_i)$ respectively.

Define $\hat{W}_i = (X_{1i}, Y_{2i}, \hat{V}_i)$, $\tilde{f}_i = \hat{f}(\hat{W}_i)$ and $\tilde{M}_i = \hat{M}(\hat{W}_i)$. Ahn (1995) shows that both $\tilde{f}_i$ and $\tilde{M}_i$ are consistent estimators of $f(W_i)$ and $M(W_i)$. The convergence rate of these two step kernel estimators depends on whether the first step estimation of $\hat{W}_i$ or the second step estimation of $\hat{f}$ and $\hat{M}$ is faster in convergence. Let $\tilde{\xi}_i = Y_i - \tilde{M}_i$, construct the test statistic as

$$\begin{aligned}\tilde{S}_1 &= \frac{1}{n}\sum_i \tilde{\xi}_i \tilde{f}_i \left[\frac{1}{(n-1)h_2^{d_2}} \sum_{j \neq i} \tilde{\xi}_j \tilde{f}_j K_2\left(\frac{X_i - X_j}{h_2}\right)\right] \\ &= \frac{1}{n(n-1)h_2^{d_2}} \sum_i \sum_{j \neq i} \left[\tilde{\xi}_i \tilde{f}_i\right]\left[\tilde{\xi}_j \tilde{f}_j\right] K_2\left(\frac{X_i - X_j}{h_2}\right),\end{aligned}$$

where $K_2 : R^{d_2} \to R$ is the kernel function and h_2 the bandwidth. The dimension of this last step kernel estimation is $d_2 = p_1 + p_2 + q$.

The above test statistic differs from that in Fan and Lee (1996) by adding the first step estimator $\hat{V}$ for the control variable. In the next we show that the control function estimator $\hat{V}$ does not affect the asymptotic distribution of the test statistic if one imposes certain conditions on the kernel function, bandwidths and the underlying distribution.

Assumption 3.2.2. *1. The density $f_1(w)$ is $r+1$ times continuously differentiable, $r \leq 2$. The density $f_2(x)$ is continuously differentiable. The density functions and their derivatives are bounded and square integrable.*

2. $\sigma^2(x) = E[(Y - E[Y|W])^2|X = x]$ is continuous and square integrable.

Assumption 3.2.3. *Kernel functions K_0, K_1 and K_2 are assumed to be product kernels with univariate functions k_0, k_1 and k_2 respectively. The univariate kernel functions satisfy the regularity conditions in below.*

1. *$\int |k_0(z)|dz < \infty$; $\int k_0(z)dz = 1$; $\int zk_0(z) = 0$;*

2. *$\int |k_1(z)|dz < \infty$; $\int k_1(z)dz = 1$; $\int z^j k_1(z) = 0$ for $1 \leq j < r$, k_1 is M dimensional differentiable;*

3. $\int |k_2(z)|dz < \infty$*;* $\int k_2(z)dz = 1$*;* $\int zk_2(z) = 0$*.*

One can see from assumption 3.2.2 that k_0, k_2 are standard bounded symmetric kernels while k_1 could be a higher order kernel if $r > 2$.

Assumption 3.2.4. *Bandwidths are restricted to satisfy:*

1. $nh_0^{d_0} \to \infty; nh_1^{d_1} \to \infty; nh_2^{d_2} \to \infty$*;*

2. $nh_1^{2r}h_2^{d_2/2} \to 0$*;* $h_2^{d_2}/h_1^{2d_1} \to 0$*;*

3. $n^{(M-1)}h_0^{d_0(M+1)}h_1^{(M+1+d_1)}h_2^{-d_2} \to \infty$*;* $nh_0^{-d_0}h_1^{2r}h_2^{d_2} \to 0$*.*

The first condition in Assumption 3.2.4 ensures that the kernel estimators involved in each step of test statistic construction are consistent. The second ensures that the asymptotic mean square error of the kernel estimator $\hat{M}_i$ and $\hat{f}_i$ is of smaller order than $n^{-1}h_2^{-d_2/2}$. These two bandwidth assumptions are also required in Fan and Lee (1996) for their test for models with exogenous covariates. The last condition in Assumption 3.2.4 is necessary so that the first step estimation for the control variable $\hat{V}_i$ does not affect the asymptotic distribution of the test statistic.

Theorem 3.2.1. *Under Assumption 3.2.1- 3.2.4, we have that under* H_0*,*

$$\tilde{\Gamma} = \frac{nh_2^{d_2/2}\tilde{S}_1}{\sqrt{2}\tilde{\sigma}_{\Gamma_1}} \to N(0,1)$$

as $n \to \infty$*, where* $\tilde{\sigma}_{\Gamma_1}$ *is the square root of*

$$\tilde{\sigma}^2_{\Gamma_1} = \frac{C}{n(n-1)h_2^{d_2}} \sum_i \sum_{i \neq j} \left[\tilde{\xi}_i \tilde{f}_i\right]^2 \left[\tilde{\xi}_j \tilde{f}_j\right]^2 K_2\left(\frac{X_i - X_j}{h_2}\right).$$

The constant $C = \int K_2^2(t)dt$ *with a* d_2 *dimensional variable* t. $\tilde{\sigma}^2_{\Gamma_1}$ *is the estimator for*

$$\sigma^2_{\Gamma_1} = CE\left[\{\xi_1 f_1(W)\}^2 E[\{\xi_1 f_1(W)\}^2|X]f(X)\right]$$

Now define the test decision rule as

$$\text{"reject } H_0^a \text{ if } \tilde{\Gamma}_1 > c_1\text{"},$$

where c_1 is the critical value derived at a certain significance level. From the convergence results discussed above, we have the following proposition that characterizes properties of the specification test. The proposition shows that the test we proposed in this section is consistent and that it has a standard normal distribution. Thus one can use the usual standard normal critical values that would be appropriate for testing against one sided alternative hypotheses eg: 1.64 for the 5% significance level and 2.33 for the 1% significance level.

Proposition 3.2.1. *Given Assumption 3.2.1-3.2.4 and that* c_1 *is a positive finite constant, we have:*

1. *Under* H_0^a, $\lim_{n\to\infty} P($*reject* $H_0^a) = 1 - \Phi(c)$,
2. *Under* H_1^a, $\lim_{n\to\infty} P($*reject* $H_0^a) = 1$,

where $\Phi(.)$ *is the cumulative distribution function of standard random normal variables.*

One should note that thus far the model has been assumed to have a separable structure. If on the other hand one had a nonseparable model such as,

$$Y = g(X, \varepsilon) = g(X_1, X_2, Y_2, \varepsilon).$$

then one can still show that $E[Y - E(Y|X_1, Y_2, V)|X] = E[E[Y - E(Y|X_1, Y_2, V)|X, V]|X] = 0$ when the null is true and $E[Y - E(Y|X_1, Y_2, V)|X] \neq 0$ when the null is false. The extension to this type of situation would require a strengthening of the assumption in (3.4) so that,

$$\varepsilon|Z, V \sim \varepsilon|V$$

3.3 Parametric and Semiparametric Specification Tests

3.3.1 Single Index Models with Endogenous Covariates

This section considers the situation where one is interested in whether or not the structural function $g(X)$ can be written in a single index form $G(X\beta)$ with the parameter vector β where both G and β are unknown. As is well known, when there are a large number of covariates a fully nonparametric model may suffer from a curse of dimensionality problem meaning that it can be difficult to obtain precise estimates given typical data availability in economic applications. The single index model represents a compromise between nonparametric and parametric specifications. In this instance the null and alternative hypotheses can be written as,

$$H_0^b : g(X) = G(X\beta) \text{ a.e. for some } \beta \in \Omega, G(.) : R \to R,$$
$$H_1^b : g(X) \neq G(X\beta) \text{ a.e. for all } \beta \in \Omega, G(.) : R \to R,$$

where Ω is a parameter set, $\Omega \in R^d$. We use the same unrestricted model setup as in the last subsection in equation (3.1)-(3.2) but suppress the regressor X_2 for simplicity. The parameter vector β is then $p_1 + q$ dimensional. Redefine in this section $d_0 = p_1 + p_3$, $d_1 = 1 + q$ and $d_2 = p_1 + q$. To define the test statistic, we replace W, given β, with $W = (X\beta, V)$ and W_i with $W_i = (X_i\beta, V_i)$. Then $\xi = Y - E(Y|W)$ and by the same argument as in last section, we know that the expression $E[\xi f(W)E[\xi f(W)|X]f(X)]$ is equal to zero when the null is true and positive when the null is false. The test statistic to be defined is the sample analog of $E[\xi f(W)E[\xi f(W)|X]f(X)]$.

Define kernel estimators of the density $f(W)$ and the conditional expectation $M(W) = E(Y|W)$ by

$$\hat{f}(W_i) = \hat{f}(X_i\beta, V_i) = \frac{1}{(n-1)h_1^{d_1}} \sum_{j \neq i} K_1 \left(\frac{(X_i - X_j)\beta}{h_1}, \frac{V_i - V_j}{h_1} \right), \quad \text{and}$$

$$\hat{M}(W_i) = \hat{M}(X_i\beta, V_i) = \frac{1}{\hat{f}(X_i\beta, V_i)} \frac{1}{(n-1)h_1^{d_1}} \sum_{j \neq i} K_1 \left(\frac{(X_i - X_j)\beta}{h_1}, \frac{V_i - V_j}{h_1} \right) Y_j.$$

Let $\hat{V}$ and $\hat{\beta}$ be the first step nonparametric estimator of V and the semiparametric up-to-scale estimator of the coefficient β; $\hat{W}_i = (X_i\hat{\beta}, \hat{V}_i)$. Refresh the definitions that $\tilde{\xi}_j = \hat{M}(\hat{W}_j)$ and $\tilde{f}_j = \hat{f}(\hat{W}_j)$. Then the test statistic for single index specification testing in models with endogenous covariates is

$$\tilde{S}_2 = \frac{1}{n(n-1)h_2^{d_2}} \sum_i \sum_{j \neq i} \left[\tilde{\xi}_i \tilde{f}_i \right] \left[\tilde{\xi}_j \tilde{f}_j \right] K_2 \left(\frac{X_i - X_j}{h_2} \right).$$

Similar to the last section, the control function V is estimated as the residual of the kernel estimation in equation (3.2), except that the set of regressor X_2 is suppressed from Z in this section for simplicity. The single index coefficient β is estimated following Blundell and Powell (2004). It is based on the matching process discussed earliest in Ahn, Ichimura and Powell (1993). The identification strategy is to build some positive semi-definite matrix such that the eigenvector corresponding to the zero eigenvalue is equal to the β coefficient up-to-scale. It is assumed that the conditions for identification are met and that one has a root n consistent estimator for the coefficient vector β.

Assumption 3.3.1. *The estimator $\hat{\beta}$ is $\sqrt{n}$ consistent, i.e. $\hat{\beta}-\beta = O_p(n^{-1/2})$.*

Under Assumption 3.2.5, the first step coefficient estimator converges faster than the kernel based estimators involved in the test statistic. Under this condition the estimation error in estimating $\hat{\beta}$ does not affect the asymptotic distribution of the test statistic and so the statistic behaves, to first order as if the value of β is actually known. Let K_0 and h_0 again denote the kernel function and bandwidth used in this first step control function estimation. To derive the asymptotic distribution of the test statistic under the null hypothesis, we assume that the kernel functions K_0, K_1 and K_2 used in this section for $\hat{S}_2$ satisfy Assumption 3.2.3 and the bandwidths h_0, h_1 and h_2 satisfy Assumption 3.2.4 with the new redefined dimension parameters d_0, d_1 and d_2.

Theorem 3.3.1. *Under Assumption 3.2.1-3.2.5, we have that under H_0^b,*

$$\tilde{\Gamma}_2 = \frac{nh_2^{d_2/2}\tilde{S}_2}{\sqrt{2}\tilde{\sigma}_{\Gamma_2}} \to N(0,1)$$

as $n \to \infty$*, where* $\tilde{\sigma}_{\Gamma_2}$ *is the square root of*

$$\tilde{\sigma}^2_{\Gamma_2} = \frac{C}{n(n-1)h_2^{d_2}} \sum_i \sum_{i \neq j} \left[\tilde{\xi}_i \tilde{f}_i\right]^2 \left[\tilde{\xi}_j \tilde{f}_j\right]^2 K_2 \left(\frac{X_i - X_j}{h_2}\right)$$

The decision rule is again to reject the null hypothesis when the estimated $\tilde{\Gamma}_2$ is larger than some critical value. From Theorem 3.3.1, we know that the single index specification test shares the asymptotic properties of the benchmark test. The critical values are distribution free and could be looked up from the normal distribution table.

3.3.2 Parametric Models with Endogenous Covariates

The testing procedures described above could also be modified to check parametric specifications in models with endogeneity. Generally, define $G(X) = (G_1(X)...G_{L_1}(X))$ with known polynomial functions $G_l(X) : R^{d_2} \to R$, $l = 1, ..., L_1$. Then given $\beta \in \Omega'^{L_1}$, $G(X)\beta = G_1(X)\beta_1 + ...G_l(X)\beta_l + ... + G_{L_1}(X)\beta_{L_1}$. For instant, if $G(X) = (X_1 \ X_1^2 \ X_1Y_2)$ and $\beta = (\beta_1 \ \beta_2 \ \beta_3)'$ then the restricted parametric model is $Y = \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1 Y_2 + u$. It is very popular in empirical studies to assume that the structural function follows the functional form $G(X)\beta$, which is polynomial in the covariate X but linear in the coefficient parameter β. Therefore it is important to be able to test whether such parametric restriction on the structural function specification is valid or not. We write the null and alternative hypotheses to be tested as:

$$H_0^c : g(X) = G(X)\beta \text{ a.e. for some } \beta \in \Omega',$$

$$H_1^c : g(X) \neq G(X)\beta \text{ a.e. for all } \beta \in \Omega'.$$

Recall from the previous subsection that $X = (X_1\ Y_2)$ and $Z = (X_1\ Z_3)$. Suppose $H(Z) = (H_1(Z)...H_{L_2}(Z))$ with known polynomial functions $H_l(X) : R^{d_0} \to R$, $l = 1, ..., L_2$. Further assume that

$$Y_2 = H(Z)\pi + V, \tag{3.8}$$

$$\varepsilon = V\rho + \eta,\ \ E(\eta|V) = 0, \tag{3.9}$$

where π is an unknown $L_2 \times 1$ parameter vector and V the error term. The error term ε in equation (3.1) could be written as a linear function of V plus an error term η. Equation (3.9) shows the origin of endogeneity of this parametric model. The control function approach is to obtain the $\sqrt{n}$ consistent coefficient estimator of the coefficient parameter vector β up-to-scale by regressing the outcome variable Y on covariates X_1, Y_2 and the control function $\hat{V}$ with $\hat{V}$ being the first step OLS residual from equation (3.8). If $G(X)$ does not include higher order polynomial terms or interaction terms of the endogenous regressor, the control function estimator is identical to the standard 2SLS estimation. If there are polynomial or interaction terms then 2SLS is invalid and the control function estimator is likely to be more efficient than the IV estimator but less robust. See Wooldridge (2007) for a detailed discussion.

Let $\hat{\beta}$ and $\hat{\rho}$ denote the control function estimators for the coefficient parameters. Now, redefine $W_i = (G(X_i)\ V_i)$, $\hat{W}_i = (G(X_i)\ \hat{V}_i)$, $\theta = (\beta\ \rho)$ and $\hat{\theta} = (\hat{\beta}\ \hat{\rho})$. Let $\xi = Y - W\theta$ and $\tilde{\xi}_i = Y_i - \hat{W}_i\hat{\theta}$. The test statistic is

$$\tilde{S}_3 = \frac{1}{n(n-1)h^{d_2}} \sum_i \sum_{j \neq i} \tilde{\xi}_i \tilde{\xi}_j K_2\left(\frac{X_i - X_j}{h_2}\right),$$

The test statistic is a sample analogue of $E[\xi E[\xi|X]f(X)]$. Note that the error term ξ is estimated parametrically under this null hypothesis. Therefore we no longer need the kernel based weighting function $\tilde{W}$ in the test statistic. The kernel function K_2 and bandwidth h_2 satisfy conditions in Assumption 3.2.3 and 3.2.4 that involve K_2 and h_2 only.

Theorem 3.3.2. *Under Assumption 3.2.1-3.2.4, we have that under H_0^c,*

$$\tilde{\Gamma}_3 = \frac{nh^{d/2}\tilde{S}_3}{\sqrt{2}\tilde{\sigma}_{\Gamma_3}} \to N(0,1)$$

as $n \to \infty$, where $\tilde{\sigma}_{\Gamma_3}$ is the square root of

$$\tilde{\sigma}^2_{\Gamma_3} = \frac{C}{n(n-1)h_2^d}\sum_i \sum_{i \neq j} \tilde{\xi}_i^2 \tilde{\xi}_j^2 K_2\left(\frac{X_i - X_j}{h_2}\right).$$

The same question about parametric structural function $g(.)$ tested against the nonparametric alternative is also discussed by Donald, Imbens and Newey (2003), Tripathi and Kitamura (2003) and Horowitz (2006). Their tests are based on the assumption in equation (3.7), which as discussed is not nested within the assumptions (3.8) and (3.9) that we use for the control function approach.

3.4 Monte Carlo Experiments

In this section we consider the small sample properties of the tests with simulated data. The data generating process is the same for each of the three tests. First, generate $X_1, Z_2 \sim i.i.d.N(0, 0.3^2)$. Error term u and v follow bivariate random normal distribution with zero mean and variance covariance matrix $(0.5\ 0.25; 0.25\ 0.5)$. $Y_2 = X_1 + Z_2 + v$ is an endogenous regressor. The outcome variable Y is generated differently depending on whether or not the null hypothesis is correct. For each of the three tests, generate the outcome variable as $Y = \Lambda + u$ when the purpose it to evaluate the test performance when the null is true and $Y = \Lambda + X_1^2/2 + u$ otherwise. In all tests conducted in this Monte Carlo section, regressors are normalized to have one standard deviation before kernel estimation is conducted. The kernel functions are product Gaussian.

In the first experiment testing for nonparametric restrictions, define $\Lambda = X_1 + Y_2$. We want to test whether X_1 belong to the structural function influencing the outcome Y after the covariate Λ is controlled, i.e. $g(\Lambda, X_1) = G(\Lambda)$. Bandwidth are chosen to be $h_0 = 0.5n^{-\frac{1}{6}}$, $h_1 = 0.5n^{-\frac{1}{6}}$ and $h_2 = cn^{-\frac{1}{2.75}}$, which satisfy the bandwidth assumptions in Section 2. The constant c is allowed to take values 0.75, 1 or 1.25 to see whether the performance of tests is sensitive to the bandwidth choice.

Table 1 reports rejection probabilities of each one-sided tests using 5% significance level and 2000 simulations. When the null hypotheses are true, the rejection probabilities are expected to get closer to 5% when the sample size gets larger. When the null hypotheses are false, the rejection probabilities should go to 1 when the sample size gets larger. We find from the first three columns of that tests generally perform as the theory predicts. The results suggest that the size of the test is not very sensitive to the choice of bandwidth. On the other hand the power of the tests does seem to depend on the choice of bandwidth.

Table 3.1: Rejection Proportions of Nonparametric and Semiparametric Tests

	Test 1			Test 2		
Sample Size	c=0.75	c=1	c=1.25	c=0.75	c=1	c=1.25
When the null is true:						
N=500	0.013	0.012	0.013	0.020	0.015	0.013
N=1000	0.016	0.017	0.013	0.024	0.019	0.017
N=2000	0.021	0.019	0.018	0.025	0.021	0.017
N=4000	0.022	0.019	0.018	0.028	0.021	0.019
When the null is false:						
N=500	0.183	0.252	0.318	0.112	0.135	0.172
N=1000	0.384	0.550	0.665	0.214	0.296	0.403
N=2000	0.729	0.886	0.953	0.443	0.625	0.767
N=4000	0.981	0.998	1.000	0.819	0.943	0.985

Then we study the small sample behavior of the single index specification test discussed in Section 3. The null hypothesis is that the structural function $g(X_1, Y_2) = G(X_1 + Y_2\beta)$ for some parameter vector β and restricted function $G(.)$. As discussed earlier, a $\sqrt{n}$ consistent coefficient estimator is needed for the test. Blundell and Powell (2004) proposed such an estimator but my simulation study found that their estimator does not have satisfactory small sample behavior. Ruthe (2008) has the same finding in his Monte Carlo experiments. For now I just use an arbitrarily constructed $\sqrt{n}$ consistent estimator that converges to the true value of $\beta = 1$. A new $\sqrt{n}$ consistent estimator is currently under study by the author that uses Ichimura's (1993) idea of Semiparametric Least Square estimation method for the semiparametric coefficient estimation in the control function approach. Bandwidths used are the same as in the last experiment. We see from the 4-6 columns of Table 1 that the single index specification test performs well with the arbitrarily constructed $\sqrt{n}$ consistent coefficient estimator.

Finally we study the small sample behavior of our parametric test for the linear specification with endogeneity. The restricted structural form under the null hypothesis is now the linear function $X_1 + Y_2\beta$ with the unknown parameter vector β. The bandwidth is chosen to be $h = cn^{-\frac{1}{6}}$ with the constant c taking values 0.4, 0.6 or 0.8. The rejection proportions from 2000 simulations are reported in Table 2. We find that the the tests perform well in small sample as the rejection proportions are well controlled around the 5% significance level when the null is true and go to 1 very quickly when the null is false. Moreover, the testing results are not very sensitive to bandwidth choices, although with very small datasets the rejection probability when the null is false is a little bit dependent on the bandwidth selection.

Table 3.2: Rejection Proportions of the Parametric Test

	Test 3		
Sample Size	c=0.4	c=0.6	c=0.8
When the null is true:			
N=100	0.035	0.026	0.018
N=200	0.036	0.033	0.029
N=400	0.037	0.030	0.026
When the null is false:			
N=100	0.206	0.337	0.454
N=200	0.508	0.777	0.915
N=400	0.909	0.997	1.000

3.5 Conclusion

This paper proposes specification tests for nonparametric models with restrictions, semiparametric single index models and parametric models with endogenous covariates. Test statistics are shown to follow standard normal distribution in the limit. Therefore, proposed tests have distribution free critical values. Both the nonparametric test and the parametric test are shown by Monte Carlo experiments to have generally good small sample properties in terms of both size and power. As for the semiparametric single index specification test, there is an issue about Blundell and Powell's (2004) up-to-scale coefficient identification strategy. The estimator does not seem to perform well in small samples, which largely affects the small sample behavior of the test. But when an arbitrary generated $\sqrt{n}$ consistent coefficient estimator series is used instead of Blundell and Powell's estimator, the test is seen to have perform well in small samples both when the null is true and when the null is false. A new identification method for the up-to-scale coefficient is now under study for the single index model with continuous endogenous covariates. The methodology is based on Ichimura's (1993) Semiparametric Least Squares approach.

One last note to make for the tests is about the convergence rate of test statistics. The rate is $nh_2^{d_2}$, where h_2 is the bandwidth of the final step kernel estimation and d_2 the dimension of covariates in the unrestricted nonparametric model. Although the convergence rate seems not to rely on the earlier steps kernel estimation of the control function and the conditional expectation $E(Y|W)$, it actually does. As we can see from assumption 3.2.4, the convergence rate $nh_2^{d_2}$ is actually constrained by dimensions of earlier step kernel estimations and characteristics of earlier step kernel functions. For example, the larger is the dimension of kernel estimations in earlier steps, the faster the bandwidth h_2 is required to converge to zero and hence the slower is the convergence rate of the test statistic. We might be able to improve the test efficiency if we construct test statistics as a function of the difference between the restricted and unrestricted average structural function estimators. This topic is out of the scope of this paper but might be an interesting topic for future researches.

Appendices

Appendix A

Chapter 1 Appendix

Proof of Theorem 1.3.1

Proof. To prove Theorem 1.3.1, it is sufficient to show that the asymptotic property holds for the $Z_1 = 0$ subsample. And all we need to show is that under Assumption 1.3.1-1.3.4,

$$[n_0 h_0^q]^{\frac{1}{2}} \left(\hat{F}_0(.|x_2) - F_0(.|x_2)\right) \Rightarrow A_1^{-1} f_0(x_2)^{-\frac{1}{2}} \mathfrak{B}\left(F_0(.|x_2)\right) \tag{A.1}$$

in $D([0, \bar{y}])$, since the theorem obviously follows from the above weak convergence result and the facts that $\hat{f}_1(x_2) \xrightarrow{p} f_1(x_2)$ and $\hat{f}_0(x_2) \xrightarrow{p} f_0(x_2)$. The proof for equation (A.1) parallels the one in Horvath and Yandell's (1988) except that multidimensional conditioning variable is allowed. First, define empirical distributions corresponding to $G_0(y, x_2)$ and $F_0(x_2)$ as follows.

$$\hat{G}_0(y, x_2) = \frac{1}{n_0} \sum_{i=1}^{n} 1(X_{1i} = 0) 1(Y_i \leq y, X_{2i} \leq x_2),$$
$$\hat{F}_0(y, x_2) = \hat{G}_0(\infty, x_2) = \frac{1}{n_0} \sum_{i=1}^{n} 1(X_{1i} = 0) 1(X_{2i} \leq x_2).$$

Then define two sequences that we will use frequently in the proof based on the empirical processes:

$$\alpha_0(y, x_2) = n_0^{\frac{1}{2}} \left(\hat{G}_0(y, x_2) - G_0(y, x_2)\right), \quad t_0(x_2) = n_0^{\frac{1}{2}} \left(\hat{F}_0(x_2) - F_0(x_2)\right)$$

In the rest of the proof, we suppress for simplicity the under-scripts and indicator functions denoting the subsample. For example, we write $\hat{G}_0(y, x_2)$ as

$\hat{G}(y, x_2) = \frac{1}{n}\sum_{i=1}^n 1(Y_i \leq y, X_{2i} \leq x_2)$, $\hat{F}_0(y, x_2)$ as $\hat{F}(y, x_2) = \frac{1}{n}\sum_{i=1}^n 1(X_{2i} \leq x_2)$, G_0, F_0 as G, F and $\alpha_0(y, x_2)$, $t_0(y, x_2)$ as $\alpha(y, x_2)$, $t(y, x_2)$.

By Neuhaus (1971) which extends Billingsley's (1999) (first edition published in 1968) weak convergence results on D([0,1]) to the space of all cadlag functions on $[0, 1]^q$, we have the following lemmas.

Lemma 1. $t \Rightarrow \gamma$, *where* γ *is a centered Gaussian process on* $[0, 1]^q$ *with covariance* $cov(\gamma(x_2), \gamma(x_2')) = F(x_2 \wedge x_2') - F(x_2)F(x_2')$ *for any* $x_2, x_2' \in \mathcal{W}$. $P[\gamma \in C] = 1$.

Lemma 2. $\alpha \Rightarrow \tau$, *where* τ *is a centered Gaussian process on* $[0, 1]^{q+1}$ *with covariance* $cov(\tau(y, x_2), \tau(y', x_2')) = G(y \wedge y', x_2 \wedge x_2') - G(y, x_2)G(y', x_2')$ *for any* $(y, x_2), (y', x_2') \in \mathcal{Y} \times \mathcal{W}$. $P[\tau \in C] = 1$.

$x_2 \wedge x_2'$ is equal to $(x_2^1 \wedge x_2'^1, ..., x_2^q \wedge x_2'^q)$, where x_2^i is the ith element of x_2, $i = 1, ..., q$. Define,

$$\begin{aligned}
\hat{g}(y, x_2) &= h^{-q} \int K\left(\frac{u - x_2}{h}\right) d_u\hat{G}(y, u) \\
&= (nh^q)^{-1} \sum_{i=1}^n 1(Y_i \leq y) \int K\left(\frac{u - x_2}{h}\right) d1(X_{2i} \leq u) \\
&= (nh^q)^{-1} \sum_{i=1}^n K\left(\frac{X_{2i} - x_2}{h}\right) 1(Y_i \leq y).
\end{aligned}$$

If not otherwise declared, integrals in the appendix will be over supports of the corresponding variables. The third equality are from integration by parts. Notice at the same time that the kernel density estimator and Nadaraya-Waston estimator we defined in Section 1.3.1 could be written as

$$\hat{f}(x_2) = \hat{g}(\infty, x_2) = h^{-q} \int K\left(\frac{u - x_2}{h}\right) d\hat{F}(u) = (nh^q)^{-1} \sum_{i=1}^n K\left(\frac{X_{2i} - x_2}{h}\right),$$

$$\hat{F}(y|x_2) = \hat{g}(y, x_2)/\hat{f}(x_2).$$

Define expectations of $\hat{g}(y, x_2)$ and $\hat{f}(x_2)$ as $\bar{g}(y, x_2)$ and $\bar{f}(x_2)$, then

$$\begin{aligned}
\bar{g}(y, x_2) &= h^{-q} \int K\left(\frac{X_{21} - x_2}{h}\right) \left(\int 1(Y_{21} \leq y) s(Y_{21}, X_{21}) dY_{21}\right) dX_{21} \\
&= h^{-q} \int K\left(\frac{X_{21} - x_2}{h}\right) g(y, X_{21}) dX_{21} \\
&= h^{-q} \int K\left(\frac{X_{21} - x_2}{h}\right) d_{X21} G(y, X_{21}) \\
&= h^{-q} \int K\left(\frac{u - x_2}{h}\right) d_u G(y, u), \\
\bar{f}(x_2) &= h^{-q} \int K\left(\frac{u - x_2}{h}\right) dF(u) = \bar{g}(\infty, x_2).
\end{aligned}$$

Also, define

$$\bar{F}(y, x_2) = \bar{g}(y, x_2)/\bar{f}(x_2).$$

From standard kernel estimation calculation (see for example Pagan and Ullah (1999)) and the uniform boundedness conditions required in Assumption 1.3.2, we have the following asymptotic results for any $x_2 \in \mathcal{W}$.

Lemma 3. *Under Assumption 1.3.1-1.3.4, we have*

1. $\hat{f}(x_2) - f(x_2) = O_p(1)$, $\bar{f}(x_2) - f(x_2) = O(h^2)$, $(nh^q)^{\frac{1}{2}}\left(\hat{f}(x_2) - \bar{f}(x_2)\right) = O_p(1)$;
2. $\sup_y |\hat{g}(y, x_2) - g(y, x_2)| = O_p(1)$, $\sup_y |\bar{g}(y, x_2) - g(y, x_2)| = O(h^2)$;
3. $\sup_y |\hat{F}(y|x_2) - F(y|x_2)| = O_p(1)$, $\sup_y |\bar{F}(y|x_2) - F(y|x_2)| = O(h^2)$.

Given the uniform convergence result of $\bar{F}(y|x_2)$, we know that $(nh^q)^{\frac{1}{2}} \sup_y \left|\left(\bar{F}(y|x_2) - F(y|x_2)\right)\right| = O\left((nh^{q+4})^{\frac{1}{2}}\right) \to 0$. Therefore, to prove Theorem 1.3.1 we only need to show the following lemma.

Lemma 4. *Under Assumption 1.3.1-1.3.3 and the first part of 1.3.4,*

$$(nh^q)^{\frac{1}{2}}\left(\hat{F}(.|x_2) - \bar{F}(.|x_2)\right) \Rightarrow \left[\frac{\int K(\phi)^2 d\phi}{f(x_2)}\right]^{\frac{1}{2}} \mathfrak{B}\left(F(.|x_2)\right)$$

Proof of Lemma 4: Define $\beta(y|x_2) = (nh^q)^{\frac{1}{2}} \left(\hat{F}(y|x_2) - \bar{F}(y|x_2)\right)$. First we want to show that

$$\sup_{y\in\mathcal{Y}} \left\{ \beta(y|x_2) - \left(\frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u-x_2}{h}\right) d_u\alpha(y,u) - \frac{h^{-\frac{q}{2}}F(y|x_2)}{f(x_2)} \int K\left(\frac{u-x_2}{h}\right) dt(u) \right) \right\}$$
$$= o_p(1). \tag{A.2}$$

Decompose $\beta(y|x)$, we get

$$\beta(y|x_2) = (nh^q)^{\frac{1}{2}} \left[\frac{\hat{g}(y,x_2)}{\hat{f}(x_2)} - \frac{\bar{g}(y,x_2)}{\bar{f}(x_2)} \right]$$
$$= (nh^q)^{\frac{1}{2}} \frac{\hat{g}(y,x_2) - \bar{g}(y,x_2)}{\bar{f}(x_2)} - (nh^q)^{\frac{1}{2}} \frac{\bar{g}(y,x_2)\left(\hat{f}(x_2) - \bar{f}(x_2)\right)}{\bar{f}^2(x_2)}$$
$$- (nh^q)^{\frac{1}{2}} \frac{(\hat{g}(y,x_2) - \bar{g}(y,x_2))\left(\hat{f}(x_2) - \bar{f}(x_2)\right)}{\bar{f}(x_2)\hat{f}(x_2)} + (nh^q)^{\frac{1}{2}} \frac{\bar{g}(y,x_2)\left(\hat{f}(x_2) - \bar{f}(x_2)\right)^2}{\bar{f}^2(x_2)\hat{f}(x_2)}$$
$$= \frac{h^{-\frac{q}{2}}}{\bar{f}(x_2)} \int K\left(\frac{u-x_2}{h}\right) d_u\alpha(y,u) - \frac{h^{-\frac{q}{2}}\bar{F}(y|x_2)}{\bar{f}(x_2)} \int K\left(\frac{u-x_2}{h}\right) dt(u)$$
$$- (nh^q)^{\frac{1}{2}} \frac{(\hat{g}(y,x_2) - \bar{g}(y,x_2))\left(\hat{f}(x_2) - \bar{f}(x_2)\right)}{\bar{f}(x_2)\hat{f}(x_2)} + (nh^q)^{\frac{1}{2}} \frac{\bar{g}(y,x_2)\left(\hat{f}(x_2) - \bar{f}(x_2)\right)^2}{\bar{f}^2(x_2)\hat{f}(x_2)} \tag{A.3}$$

The first two terms in the RHS of equation (A.3) converges (uniformly w.r.t y) to $\frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u-x_2}{h}\right) d_u\alpha(y,u)$ and $\frac{h^{-\frac{q}{2}}F(y|x_2)}{f(x_2)} \int K\left(\frac{u-x_2}{h}\right) dt(u)$ respectively since $\bar{f}(x_2) - f(x_2) = o(1)$ and $\sup_y |\bar{F}(y|x_2) - F(y|x_2)| = o(1)$ under Lemma 3.

The third and fourth terms in the RHS of equation (A.3) go to zero uniformly because $\hat{f}(x_2) - \bar{f}(x_2) = O_p\left((nh^q)^{-\frac{1}{2}}\right)$ and that

$$\sup_y |\hat{g}(y,x_2) - \bar{g}(y,x_2)| = (nh^q)^{-\frac{1}{2}} \sup_y |h^{-\frac{q}{2}} \int K\left(\frac{u-x_2}{h}\right) d_u\alpha(y,u)| \tag{A.4}$$
$$= O_p\left((nh^q)^{-\frac{1}{2}}\right), \tag{A.5}$$

The second equality will be shown in a moment in (A.6). The result in (A.2) hence follows.

Let H be the distribution function of y and I^j the marginal distribution function of X_2^j, $j = 1, 2, ..., p$. Define Copula function of cumulative distribution function $G(y, x_2) = J(H(y), I(x_2))$, where $I(x_2) = (I^1(x_2^1), ..., I^q(x_2^q))$. Then the Gaussian process τ in Lemma 2 satisfies

$$\{\tau(y, x_2),\ (y, x_2) \in \mathcal{Y} \times \mathcal{W}\} \overset{D}{=} \{W_J(H(y), I(x_2)) - G(y, x_2) W_J(1, \iota),\ (y, x_2) \in \mathcal{Y} \times \mathcal{W}\}$$

where ι is a vector of q ones, W_J is a q+1 dimensional Wiener process with $E[W_J(s,t)] = 0$ and $E[W_J(s,t)W_J(s',t')] = J(s \wedge s', t \wedge t')$, for $(s,t), (s',t') \in \mathcal{Y} \times \mathcal{W}$. $\overset{d}{=}$ means that two random variables (processes) have the same distribution.

Lemma 2 gives that when $n \to \infty$,

$$\left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u \alpha(y, u), y \in \mathcal{Y} \right\} \overset{d}{=} \left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u \tau(y, u), y \in \mathcal{Y} \right\}$$
$$\overset{d}{=} \left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u W_J\left(H(y), I(u)\right) - \frac{h^{-\frac{q}{2}}}{f(x_2)} W_J(1, \iota) \int K\left(\frac{u - x_2}{h}\right) g(y, u) du, y \in \mathcal{Y} \right\}$$

Since $\sup_y \left| \int K\left(\frac{u - x_2}{h}\right) g(y, u) du \right| \leq h^q \sup_{y,u} |g(y, u)| \int |K(\phi)| d\phi = O\left(h^q\right)$ under Assumptions 1.3.2 and 1.3.3, we have that

$$\left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u \alpha(y, u), y \in \mathcal{Y} \right\} \overset{d}{=} \left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u W_J(H(y), I(u)), y \in \mathcal{Y} \right\}.$$

Calculations show,

$$E\left(h^{-\frac{q}{2}} \int K\left(\frac{u - x_2}{h}\right) d_u W_J(H(y), I(u)) \right) = 0$$
$$E\left(h^{-q} \int K\left(\frac{u - x_2}{h}\right) d_u W_J(H(y), I(u)) \int K\left(\frac{u - x_2}{h}\right) d_u W_J(H(y'), I(u)) \right)$$
$$= h^{-q} \int K\left(\frac{u - x_2}{h}\right)^2 g(y \wedge y', u) du \overset{\text{let}}{=} \bar{l}(y \wedge y')$$

Then, we know from the Gaussian Characterization of the Brownian Motion that

$$\left\{ \frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u \alpha(y, u), y \in \mathcal{Y} \right\} \overset{d}{=} \left\{ W\left(\bar{l}(y)\right), y \in \mathcal{Y} \right\}$$

where W is a standard Wiener process.

Let $l(y) = g(y, x_2) \int K(\phi)^2 d\phi$, we know that $\sup_y |\bar{l}(y) - l(y)| = O(h)$ under Assumption 1.3.2, 1.3.3 and 1.3.4. By Theorem 1.1.1 in Csorgo and Revesz (1981), we know that

$$\sup_y |W(\bar{l}(y)) - W(l(y))| \overset{a.s.}{=} O(2hlog\frac{1}{h}) \overset{a.s.}{\to} 0$$

Together with the rescaling property of Brownian Motion, we have that $\forall x_2 \in \mathcal{W}$,

$$\frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) d_u\alpha(., u) \Rightarrow \left[\frac{\int K(\phi)^2 d\phi}{f(x_2)}\right]^{\frac{1}{2}} W(F(.|x_2)). \tag{A.6}$$

Likewise $\forall x_2 \in \mathcal{W}$,

$$\frac{h^{-\frac{q}{2}}}{f(x_2)} \int K\left(\frac{u - x_2}{h}\right) dt(u) \Rightarrow \left[\frac{\int K(\phi)^2 d\phi}{f(x_2)}\right]^{\frac{1}{2}} W(F(\infty|x_2)).$$

It is obvious that $F(\infty|x_2) = 1$. Therefore, we finally have that

$$\begin{aligned}\beta(.|x_2) &\Rightarrow \left[\frac{\int K(\phi)^2 d\phi}{f(x_2)}\right]^{\frac{1}{2}} [W(F(.|x_2)) - F(.|x_2)W(1)] \\ &\overset{d}{=} \left[\frac{\int K(\phi)^2 d\phi}{f(x_2)}\right]^{\frac{1}{2}} \mathcal{B}(F(.|x_2)), \quad \forall x_2 \in \mathcal{W}.\end{aligned}$$

□

Proof of Proposition 1.3.1

Proof. Firstly we want to show part 1 of the proposition, the asymptotic property of the test statistic when the null hypothesis is true. For all $x_2 \in \mathcal{W}$, define

$$\hat{T}_1(.|x_2) = A_1 \left(\frac{n_1 h_1^q \hat{f}_1(x_2) n_0 h_0^q \hat{f}_0(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)}\right)^{\frac{1}{2}} \left[\left(\hat{F}_1(.|x_2) - F_1(.|x_2)\right) - \left(\hat{F}_0(.|x_2) - F_0(.|x_2)\right)\right].$$

The test statistic satisfies

$$\begin{aligned}\hat{S}_1(x_2) &\leq \sup_{y \in \mathcal{Y}} \hat{T}_1(y|x_2) + A_1 \left(\frac{n_1 h_1^q \hat{f}_1(x_2) n_0 h_0^q \hat{f}_0(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)}\right)^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} (F_1(y|x_2) - F_0(y|x_2)) \\ &\leq \sup_{y \in \mathcal{Y}} \hat{T}_1(y|x_2).\end{aligned}$$

with equality holds when $F_1(.|x_2) = F_0(.|x_2)$. Let bandwidths be $h_0 = \delta_0 n_0^{-\frac{1}{\Delta_0}}$ and $h_1 = \delta_1 n_1^{-\frac{1}{\Delta_1}}$. If $\Delta_0 > \Delta_1$, $n_0 h_0^q$ goes to infinity faster than $n_1 h_1^q$. Together with Theorem 1.3.1,

$$\begin{aligned}
\hat{T}_1(.|x_2) &= \left(\frac{n_0 h_0^q \hat{f}_0(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)}\right)^{\frac{1}{2}} A_1 \left[n_1 h_1^q \hat{f}_1(x_2)\right]^{\frac{1}{2}} \left(\hat{F}_1(.|x_2) - F_1(.|x_2)\right) \\
&\quad - \left(\frac{n_1 h_1^q \hat{f}_1(x_2)}{n_1 h_1^q \hat{f}_1(x_2) + n_0 h_0^q \hat{f}_0(x_2)}\right)^{\frac{1}{2}} A_1 \left[n_0 h_0^q \hat{f}_0(x_2)\right]^{\frac{1}{2}} \left(\hat{F}_0(.|x_2) - F_0(.|x_2)\right) \\
&\quad \to A_1 \left[n_1 h_1^q \hat{f}_1(x_2)\right]^{\frac{1}{2}} \left(\hat{F}_1(.|x_2) - F_1(.|x_2)\right) \\
&\quad \Rightarrow \mathfrak{B}\left(F_1(.|x_2)\right).
\end{aligned}$$

By the Continuous Mapping Theorem,

$$\lim_{n\to\infty} P(\sup_{y\in\mathcal{Y}} \hat{T}_1(x_2 > c_1)) = P\left(\sup_{y\in\mathcal{Y}} \mathfrak{B}\left(F_1(y|x_2)\right) > c_1\right) = P\left(\sup_{t} \mathfrak{B}(t) > c_1\right).$$

Therefore, we get the inequality in the first part of Proposition 1.3.1 in the situation bandwidth h_0 goes to zero slower than h_1. Equality in Proposition 1.3.1.1 holds when $\hat{S}_1(x_2) = \sup_{y\in\mathcal{Y}} \hat{T}_1(x_2)$, i.e. $F_1(.|x_2) = F_0(.|x_2)$. Likewise, if $\Delta_0 < \Delta_1$, $\hat{T}_1(x_2)$ converges weakly to $\mathfrak{B}\left(F_0(.|x_2)\right)$ and we also get the results stated in Proposition 1.3.1.1.

Let $\lambda_n = \frac{n_1}{n}$, $\lambda = \lim_{n\to\infty} \lambda_n = P(X_1 = 1) \in (0,1)$. If $\Delta_1 = \Delta_2$, $\frac{n_1 h_1^q \hat{f}_1(x_2)}{n_0 h_0^q \hat{f}_0(x_2)} \xrightarrow{p} \left(\frac{\delta_1}{\delta_2}\right)^q \left(\frac{\lambda}{1-\lambda}\right)^{1-\frac{q}{\Delta_1}} \frac{f_1(x_2)}{f_0(x_2)} \overset{let}{\equiv} \xi$, $\xi > 0$. Then $\hat{T}_1(x_2)$ converges weakly to a linear combination of two changed time random Brownian Bridge processes:

$$\begin{aligned}
\hat{T}_1(.|x_2) &\Rightarrow \sqrt{\frac{1}{\xi+1}}\mathfrak{B}\left(F_1(.|x_2)\right) - \sqrt{\frac{\xi}{\xi+1}}\mathfrak{B}\left(F_0(.|x_2)\right) \\
&\overset{let}{\equiv} \bar{T}_1(.|x_2).
\end{aligned}$$

Denote the limiting random variable $\bar{T}_1(.|x_2)$ as $\bar{T}_1^0(.|x_2)$ when $F_1(.|x_2) = F_0(.|x_2)$, $\bar{T}_1^0(.|x_2) \overset{d}{=} \mathfrak{B}(F_0(.|x_2))$. Let $\mathcal{Y}^*$ denote the set of y values for which $F_1(y|x_2) = F_0(y|x_2)$. Then all we need to show for Proposition 1.3.1.1 in the situation where $\Delta_0 = \Delta_1$ is that

$$P(\hat{S}_1 > c_1) \to P\left(\sup_{y\in\mathcal{Y}^*} \bar{T}_1(y|x_2) > c_1\right) \tag{A.7}$$

and that

$$P\left(\sup_{y\in\mathcal{Y}^*}\bar{T}_1(y|x_2)>c_1\right)\leq P\left(\sup_{y\in\mathcal{Y}^*}\bar{T}_1^0(y|x_2)>c_1\right)\leq P\left(\sup_{y\in\mathcal{Y}}\bar{T}_1^0(y|x_2)>c_1\right), \tag{A.8}$$

with equality holds when $F_1(.|x_2)=F_0(.|x_2)$. Proofs for equation (A.7) and (A.8) are similar to those provided for equation (18) and (26) in Barrett and Donald (2003) and are omitted here. The tightness condition obtained from the weak convergence result in Theorem 1.3.1 is the workhorse of proof for equation (A.8).

To prove the second part of the proposition, we note that if the alternative hypothesis is true, there exists some $y^+\in\mathcal{Y}$ such that

$$F_1\left(y^+|x\right)-F_0\left(y^+|x\right)=\eta>0.$$

Then we know

$$\hat{S}_1(x_2)\geq\hat{T}_1\left(y^+|x_2\right)+A_1\left(\frac{n_1h_1^q\hat{f}_1(x_2)n_0h_0^q\hat{f}_0(x_2)}{n_1h_1^q\hat{f}_1(x_2)+n_0h_0^q\hat{f}_0(x_2)}\right)^{\frac{1}{2}}\left(F_1\left(y^+|x_2\right)-F_0\left(y^+|x_2\right)\right)$$
$$\to\infty$$

The first inequality holds because $y^+\in\mathcal{Y}$; the second holds from Theorem 1.3.1 and Assumption 1.3.4. The result in the second part of Proposition 1.3.1 hence follows. □

Proof of Theorem 1.3.2

Proof. This proof is carried out using the same idea as in Theorem 1.3.1. As long as we show that under Assumption 1.3.1 and 1.3.5-1.3.7,

$$\left[nh^{p+2}\right]^{\frac{1}{2}}\left(\hat{F}^{(1)}(.|x)-F^{(1)}(.|x)\right)\Rightarrow A_2^{-1}f(x)^{-\frac{1}{2}}\mathfrak{B}(F(.|x)) \tag{A.9}$$

in $D([0,\bar{y}])$, we get the weak convergence result stated in Theorem 1.3.2 as $\hat{f}(x)\xrightarrow{p}f(x)$ under the assumptions. Now define empirical distributions corresponding to $G(y,x)$ and $F(x)$ and two sequences based on the processes:

$$\hat{G}(y,x)=\frac{1}{n}\sum_{i=1}^n 1(Y_i\leq y,X_i\leq x);\quad \hat{F}(y,x)=\hat{G}(\infty,x)=\frac{1}{n}\sum_{i=1}^n 1(X_i\leq x).$$
$$\alpha(y,x)=n^{\frac{1}{2}}\left(\hat{G}(y,x)-G(y,x)\right);\quad t(x)=n^{\frac{1}{2}}\left(\hat{F}(x)-F(x)\right)$$

Notice that $\hat{g}^{(1)}(y,x)$ and $\hat{f}^{(1)}(x)$ defined in Section 1.3.2 are equal to

$$\hat{g}^{(1)}(y,x) = -h^{-(p+1)} \int K_1^{(1)}\left(\frac{u-x}{h}\right) d_u \hat{G}(y,u),$$
$$\hat{f}^{(1)}(x) = \hat{g}^{(1)}(\infty,x) = -h^{-(p+1)} \int K_1^{(1)}\left(\frac{u-x}{h}\right) d_u \hat{F}(u).$$

While their expectations are

$$\bar{g}^{(1)}(y,x) = E[\hat{g}^{(1)}(y,x)] = -h^{-(p+1)} \int K_1^{(1)}\left(\frac{u-x}{h}\right) d_u G(y,u)$$
$$\bar{f}^{(1)}(x) = E[\hat{f}^{(1)}(x)] = -h^{-(p+1)} \int K_1^{(1)}\left(\frac{u-x}{h}\right) d_u F(u) = \bar{g}^{(1)}(\infty,x)$$

Define

$$\bar{F}^{(1)}(y|x) = \bar{f}(x)^{-1}\left[\bar{g}^{(1)}(y,x) - \bar{F}(y|x)\bar{f}^{(1)}(x)\right]$$

Call the partial derivative of $f(x)$ and $g(y,x)$ with respect to x_1 $f^{(1)}(x)$ and $g^{(1)}(y,x)$. From standard kernel estimation calculation and the uniform boundedness conditions in Assumption 1.3.5, we know that for any $x \in \mathcal{X}$, we have the following asymptotic results.

Lemma 5. *Under Assumption 1.3.1, 1.3.5-1.3.7, we have*

1. $\hat{f}^{(1)}(x) - f^{(1)}(x) = O_p(1)$, $\bar{f}^{(1)}(x_2) - f(x_2) = O(h^2)$, $(nh^q)^{\frac{1}{2}}\left(\hat{f}^{(1)}(x_2) - \bar{f}^{(1)}(x_2)\right) = O_p(1)$;
2. $\sup_y |\hat{g}^{(1)}(y,x_2) - g^{(1)}(y,x_2)| = O_p(1)$, $\sup_y |\bar{g}^{(1)}(y,x_2) - g^{(1)}(y,x_2)| = O(h^2)$;
3. $\sup_y |\hat{F}^{(1)}(y|x_2) - F^{(1)}(y|x_2)| = O_p(1)$, $\sup_y |\bar{F}^{(1)}(y|x_2) - F^{(1)}(y|x_2)| = O(h^2)$;

Again given the uniform convergence result of $\bar{F}^{(1)}(y|x_2)$, we know that $(nh^{p+2})^{\frac{1}{2}} \sup_y \left|\left(\bar{F}^{(1)}(y|x_2) - F^{(1)}(y|x_2)\right)\right| = O\left((nh^{p+6})^{\frac{1}{2}}\right) \to 0$. Therefore, to prove Theorem 1.3.2 we only need to show the following lemma.

Lemma 6. *Under Assumption 1.3.1, 1.3.5-1.3.6 and the first part of 1.3.7,*

$$(nh^{p+2})^{\frac{1}{2}}\left(\hat{F}^{(1)}(.|x) - \bar{F}^{(1)}(.|x)\right) \Rightarrow \left[\frac{\int K_1^{(1)}(\phi)^2 d\phi}{f(x)}\right]^{\frac{1}{2}} \mathfrak{B}(F(.|x))$$

Proof of Lemma 6: Define $r(y|x) = (nh^{p+2})^{\frac{1}{2}} \left(\hat{F}^{(1)}(y|x) - \bar{F}^{(1)}(y|x)\right)$. First we show that

$$\sup_{y\in\mathcal{Y}} \left\{ r(y|x) - \left(\frac{h^{-\frac{p}{2}}}{f(x)} \int -K_1^{(1)} \left(\frac{u-x}{h}\right) d_u\alpha(y,u) - \frac{h^{-\frac{p}{2}}F(y|x)}{f(x)} \int -K_1^{(1)} \left(\frac{u-x}{h}\right) dt(u) \right) \right\}$$
$$= o_p(1). \tag{A.10}$$

Decompose $r(y|x)$, we get

$$\begin{aligned} r(y|x) &= (nh^{p+2})^{\frac{1}{2}} \left[\frac{\hat{g}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{g}^{(1)}(y,x)}{\bar{f}(x)} - \left(\frac{\hat{f}^{(1)}(x)}{\hat{f}(x)}\hat{F}(y|x) - \frac{\bar{f}^{(1)}(x)}{\bar{f}(x)}\bar{F}(y|x) \right) \right] \\ &= (nh^{p+2})^{\frac{1}{2}} \left(\frac{\hat{g}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{g}^{(1)}(y,x)}{\bar{f}(x)} \right) - (nh^{p+2})^{\frac{1}{2}} \bar{F}(y|x) \left(\frac{\hat{f}^{(1)}(x)}{\hat{f}(x)} - \frac{\bar{f}^{(1)}(x)}{\bar{f}(x)} \right) \\ &- (nh^{p+2})^{\frac{1}{2}} \frac{\hat{f}^{(1)}(x)\left(\hat{F}(y|x) - \bar{F}(y|x)\right)}{\hat{f}(x)} \\ &= (nh^{p+2})^{\frac{1}{2}} \left(\frac{\hat{g}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{g}^{(1)}(y,x)}{\bar{f}(x)} \right) - (nh^{p+2})^{\frac{1}{2}} \bar{F}(y|x) \left(\frac{\hat{f}^{(1)}(x)}{\hat{f}(x)} - \frac{\bar{f}^{(1)}(x)}{\bar{f}(x)} \right) \\ &+ O_p(1) \end{aligned} \tag{A.11}$$

The third equality is due to the facts that $\forall x \in \mathcal{X}$ $\hat{f}^{(1)}(x)/\hat{f}(x)$ converges in probability to some constant bounded above/away from zero and that

$$\sup_{y\in\mathcal{Y}} \left| (nh^p)^{\frac{1}{2}} \left(\hat{F}(y|x) - \bar{F}(y|x)\right) \right| = O_p(1)$$

as Assumption 1.3.5-1.3.7.1 are stronger than Assumption 1.3.2-1.3.4.1.

Now we consider the first term in the RHS of equation (A.11).

$$
\begin{aligned}
&(nh^{p+2})^{\frac{1}{2}} \left(\frac{\hat{g}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{g}^{(1)}(y,x)}{\bar{f}(x)} \right) \\
&= (nh^{p+2})^{\frac{1}{2}} \frac{\hat{g}^{(1)}(y,x) - \bar{g}^{(1)}(y,x)}{\bar{f}(x)} - (nh^{p+2})^{\frac{1}{2}} \frac{\bar{g}^{(1)}(y,x)\left(\hat{f}(x) - \bar{f}(x)\right)}{\bar{f}^2(x)} \\
&- (nh^{p+2})^{\frac{1}{2}} \frac{\left(\hat{g}^{(1)}(y,x) - \bar{g}^{(1)}(y,x)\right)\left(\hat{f}(x) - \bar{f}(x)\right)}{\bar{f}(x)\hat{f}(x)} + (nh^{p+2})^{\frac{1}{2}} \frac{\bar{g}^{(1)}(y,x)\left(\hat{f}(x) - \bar{f}(x)\right)^2}{\bar{f}^2(x)\hat{f}(x)} \\
&= h^{-\frac{p}{2}} \int -K_1^{(1)} \left(\frac{u-x}{h} \right) d_u\alpha(y,u) - (nh^{p+2})^{\frac{1}{2}} \frac{\bar{g}^{(1)}(y,x)\left(\hat{f}(x) - \bar{f}(x)\right)}{\bar{f}^2(x)} \\
&- (nh^{p+2})^{\frac{1}{2}} \frac{\left(\hat{g}^{(1)}(y,x) - \bar{g}^{(1)}(y,x)\right)\left(\hat{f}(x) - \bar{f}(x)\right)}{\bar{f}(x)\hat{f}(x)} + (nh^{p+2})^{\frac{1}{2}} \frac{\bar{g}^{(1)}(y,x)\left(\hat{f}(x) - \bar{f}(x)\right)^2}{\bar{f}^2(x)\hat{f}(x)}
\end{aligned}
\tag{A.12}
$$

The last three parts in the RHS of equation (A.12) could be shown uniformly converging to zero by results in Lemma 3, Lemma 5 and the fact that $\sup_y \left|\hat{g}^{(1)}(y,x) - \bar{g}^{(1)}(y,x)\right| = o_p(1)$ (which could be shown in the same way as showing equation (A.4) in the proof for Theorem 1.3.1.). Therefore, we have that when $n \to \infty$,

$$
\sup_{y\in\mathcal{Y}} \left\{ (nh^{p+2})^{\frac{1}{2}} \left(\frac{\hat{g}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{g}^{(1)}(y,x)}{\bar{f}(x)} \right) - \frac{h^{-\frac{p}{2}}}{f(x)} \int -K_1^{(1)} \left(\frac{u-x}{h} \right) d_u\alpha(y,u) \right\} = O_p(1).
$$

Likewise, the second term in the RHS of equation (A.11) satisfy that as $n \to \infty$,

$$
\sup_{y\in\mathcal{Y}} \left\{ (nh^{p+2})^{\frac{1}{2}} \left(\frac{\hat{f}^{(1)}(y,x)}{\hat{f}(x)} - \frac{\bar{f}^{(1)}(y,x)}{\bar{f}(x)} \right) \bar{F}(y|x) - \frac{h^{-\frac{p}{2}}F(y|x)}{f(x)} \int -K_1^{(1)} \left(\frac{u-x}{h} \right) dt(u) \right\}
$$
$$
= O_p(1).
$$

Adding all the pieces together gives us equation (A.10). Lemma 6 is then proved by using the same calculations as in the proof for Lemma 4 after showing equation (A.2). □

Proof of Proposition 1.3.2

Proof. To prove part 1 of the proposition, we note that

$$\begin{aligned}\hat{S}_2 &\leq A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\sup_{y\in\mathcal{Y}}\left(\hat{F}^{(1)}(y|x)-F^{(1)}(y|x)\right)+A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\sup_{y\in\mathcal{Y}}F^{(1)}(y|x)\\ &\leq A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\sup_{y\in\mathcal{Y}}\left(\hat{F}^{(1)}(y|x)-F^{(1)}(y|x)\right)\end{aligned}$$

The first inequality is due to the property of supremum and the second uses the fact that under $H^2_{0,x}$, $F^{(1)}(.|x)\leq 0$. Both inequalities turn to equality when $F^{(1)}(.|x)=0$. Then, the inequality in Proposition 1.3.2.1 directly follows from Theorem 1.3.2 and the Continuous Mapping Theorem. Equality in Proposition 1.3.2.1 holds when $F^{(1)}(.|x)=0$.

To prove the second part, we note that if the alternative hypothesis is true, there exists some $y^+\in\mathcal{Y}$ such that

$$F^{(1)}\left(y^+|x\right)=\eta>0$$

. Then we know

$$\begin{aligned}\hat{S}_2 &\geq A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\hat{F}^{(1)}\left(y^+|x\right)\\ &\geq A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}\left(\hat{F}^{(1)}\left(y^+|x\right)-F^{(1)}\left(y^+|x\right)\right)+A_2\left[nh^{p+2}\hat{f}(x)\right]^{\frac{1}{2}}F^{(1)}\left(y^+|x\right)\\ &\to\infty\end{aligned}$$

The first inequality holds because $y^+\in\mathcal{Y}$; the second holds from Theorem 1.3.2 and Assumption 1.3.7. The result in the second part of Proposition 1.3.2 hence follows. □

Proof of Corollary 1.4.1

Proof. Here we prove the corollary for the dummy X_1 covariate case. The corollary for continuous X_1 covariates could be shown parallel to the steps below.

The proof has two steps. The first step is to show the following lemma. The second step uses the same arguments as in the proof for Proposition 1.3.1 to show the corollary from the following lemma and is omitted here.

Lemma 7. *Under Assumption 1.3.1-1.3.3, 1.4.1, 1.4.2 and 1.4.3,*

$$A_1[n_0h_0^q\hat{f}_0(x_2\hat{\theta})]^{\frac{1}{2}}(\hat{F}_0(.|x_2\hat{\theta})-\tilde{F}_0(.|x_2\theta))\Rightarrow\mathfrak{B}(\tilde{F}_0(.|x_2))$$

in $D([0,\bar{y}])$

A corresponding result also holds for the subsample with $X_1 = 1$. To show the lemma, we only need to show the following equation holds as Lemma 7 then follows from applying Theorem 1.3.1 to the case of single index conditioning variable $x_2\theta_2$.

$$\sup_y \left(\hat{F}_0(y|x_2\hat{\theta}_2) - \hat{F}_0(y|x_2\theta_2) \right) \to o_p([n_0h_0]^{-\frac{1}{2}}), \tag{A.13}$$

Using the same notations as in the proof for Theorem 1.3.1 and suppressing the under-scripts and indicator functions denoting the subsample, we write that

$$\begin{aligned}
\hat{F}(y|x_2\hat{\theta}_2) - \hat{F}(y|x_2\theta_2) &= \frac{\hat{g}(y, x_2\hat{\theta}_2)}{\hat{f}(x_2\hat{\theta}_2)} - \frac{\hat{g}(y, x_2\theta_2)}{\hat{f}(x_2\theta_2)} \\
&= \frac{\hat{g}(y, x_2\hat{\theta}_2) - \hat{g}(y, x_2\theta_2)}{\hat{f}(x_2\theta_2)} - \frac{\hat{g}(x_2\hat{\theta}_2)\left(\hat{f}(x_2\hat{\theta}_2) - \hat{f}(x_2\theta_2)\right)}{\hat{f}(x_2\hat{\theta}_2)\hat{f}(x_2\theta_2)}.
\end{aligned}$$

Since $\hat{g}(y, x_2\hat{\theta}_2)$ is uniformly bounded and both $\hat{f}(x_2\theta_2)$ and $\hat{f}(x_2\hat{\theta}_2)$ are bounded away from zero, we only need to show that $\sup_y \left(\hat{g}(y, x_2\hat{\theta}_2) - \hat{g}(y, x_2\theta_2) \right) = o_p([nh]^{-\frac{1}{2}})$ and $\hat{f}(x_2\hat{\theta}_2) - \hat{f}(x_2\theta_2) = o_p([nh]^{-\frac{1}{2}})$. Let ϕ_i be any number between $[(X_{2i} - x_2)\theta]/h$ and $[(X_{2i} - x_2)\hat{\theta}]/h$. By Taylor expansion, we know that

$$\begin{aligned}
\hat{f}(x_2\hat{\theta}_2) - \hat{f}(x_2\theta_2) &= \sum_{i=1}^{n} \frac{1}{nh} K\left(\frac{(X_{2i} - x_2)\hat{\theta}}{h}\right) - \sum_{i=1}^{n} \frac{1}{nh} K\left(\frac{(X_{2i} - x_2)\theta}{h}\right) \\
&= \sum_{i=1}^{n} \frac{1}{nh} \sum_{s=1}^{M} \frac{1}{s!} K^{(s)}\left(\frac{(X_{2i} - x_2)\theta}{h}\right) \left(\frac{X_{2i} - x_2}{h}\right)^s \left(\hat{\theta} - \theta\right)^s \\
&+ \sum_{i=1}^{n} \frac{1}{nh} \frac{1}{M+1} K^{(M+1)}(\phi_i) \left(\frac{X_{2i} - x_2}{h}\right)^{M+1} \left(\hat{\theta} - \theta\right)^{M+1} \\
&= \sum_{s=1}^{M} I^s + I^{M+1}.
\end{aligned}$$

Now the task is to show that $I^s = o_p([nh]^{-\frac{1}{2}})$, for all $s = 1, ..., M+1$. First,

$$\begin{aligned}
E[I^1] &= \int \frac{1}{h} K^{(1)}\left(\frac{(X_{21} - x_2)\theta}{h}\right) \frac{X_{21} - x_2}{h} \left(\hat{\theta} - \theta\right) f(X_{21}\theta) dX_{21}\theta \\
&= \int K^{(1)}(\xi) \frac{\xi}{\theta} \left(\hat{\theta} - \theta\right) f(x_{21}\theta + \xi h) d\xi \\
&= O_p(n^{-\frac{1}{2}}) = o_p([nh]^{-\frac{1}{2}})
\end{aligned}$$

Likewise, $E[(I^1)^2] = o_p([nh]^{-1})$. Hence, we get that $I^1 = o_p([nh]^{-\frac{1}{2}})$. Similarly, we obtain that $I^s = o_p([nh]^{-\frac{1}{2}})$, for $s = 2, ..., M$. Meanwhile, under the additional bandwidth condition in Assumption 1.4.3 the last residual term in the Taylor expansion satisfies that

$$
\begin{aligned}
E[I^{M+1}] &= \int \frac{1}{h} \frac{1}{M+1} K^{(M+1)}(\phi_1) \left(\frac{X_{21} - x_2}{h} \right)^{M+1} \left(\hat{\theta} - \theta \right)^{M+1} f(X_{21}) dX_{21} \\
&= O_p(n^{-M/2-1/2} h^{-M-2}) = o_p([nh]^{-\frac{1}{2}})
\end{aligned}
$$

and $E[(I^{M+1})^2] = o_p([nh]^{-1})$. Hence $I^{M+1} = o_p([nh]^{-1})$.

Similarly, we could show that $\sup_y \left(\hat{g}(y, x_2\hat{\theta}_2) - \hat{g}(y, x_2\theta_2) \right) = o_p([nh]^{-\frac{1}{2}})$ and get the results in equation A.13 and Lemma 7.

$\square$

Appendix B

Chapter 3 Appendix

Proof of Theorem 3.2.1

Proof. The proof borrows from the proof idea for Theorem 4.2 in Fan and Lee (1996) and shows that the first step nonparametric estimator for V would not affect the asymptotic properties of the test once bandwidths are required such that the convergence rate in the first step is somewhat faster than the unrestricted estimator in the second step test statistic.

Denote $\hat{\xi}_i = Y_i - \hat{M}(X_{1i}, Y_{2i}, V_i)$, the residual from kernel estimation of equation (3.2) with known control variable for sample observation i. From the definition of nonparametric estimators $\hat{M}$ and $\hat{f}$, we know that

$$
\begin{aligned}
\tilde{\xi}_i \tilde{f}_i &= \frac{1}{(n-1)h_1^{d_1}} \sum_{i \neq j} (Y_i - Y_j) K_{1ij}^{\hat{w}} \\
&= \hat{\xi}_i \hat{f}_i + \frac{1}{(n-1)h_1^{d_1}} \sum_{i \neq j} (Y_i - Y_j) \left[K_{1ij}^{\hat{w}} - K_{1ij}^{w} \right],
\end{aligned}
$$

where $K_{1ij}^{\hat{w}}$ is defined as the abbreviation for $K_1\left(\frac{\hat{W}_i - \hat{W}_j}{h_1}\right)$ and K_{1ij}^{w} for $K_1\left(\frac{W_i - W_j}{h_1}\right)$. Then the test statistic

$$
\begin{aligned}
\tilde{S}_1 &= \frac{1}{n(n-1)h_2^{d_2}} \sum_i \sum_{j \neq i} \left[\hat{\xi}_i \hat{f}_i\right] \left[\hat{\xi}_j \hat{f}_j\right] K_{1ij}^{w} \\
&+ \frac{1}{n(n-1)^2 h_1^{d_1} h_2^{d_2}} \sum_i \sum_{j \neq i} \sum_{l \neq j} \left[\hat{\xi}_i \hat{f}_i\right] (Y_j - Y_l) \left[K_{jl}^{\hat{w}} - K_{jl}^{w}\right] K_{2ij}^{w} \\
&+ \frac{1}{n(n-1)^3 h_1^{2d_1} h_2^{d_2}} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq i} (Y_i - Y_k)(Y_j - Y_l) \left[K_{ik}^{\hat{w}} - K_{ik}^{w}\right] \left[K_{jl}^{\hat{w}} - K_{jl}^{w}\right] K_{2ij}^{w} \\
&\stackrel{def}{=} I_1 + I_2 + I_3.
\end{aligned}
$$

The asymptotic distribution of I_1 is studied in Fan and Lee (1996).

Lemma 8. *Under Assumption 3.2.1-3.2.4 and H_0^a,*

$$\frac{nh_2^{d_2/2} I_1}{\sqrt{2}\sigma_{\Gamma}} \to N(0,1).$$

Next, we want to show that both I_2 and I_3 are $o_p(n^{-1}h_2^{-d_2/2})$ under the smoothness assumptions on the joint distribution of variables, the kernel assumptions and the bandwidths assumptions in 3.2.1-3.2.4. For simplicity of the Taylor expansion series, we prove the result assuming that the endogenous variable Y_2 and hence the control variable V is single dimensional, i.e., $q = 1$. The multi-dimensional case with a fixed number of dimension q for the endogenous variable follows readily. Define $K_{1jl}^o = K_1\left(\frac{X_{1j}-X_{1l}}{h_1}, \frac{Y_{2j}-Y_{2l}}{h_1}\right)$. Because K_1 is product kernel, $K_{1jl} = K_{1jl}^o k((V_j - V_l)/h_1)$. Then by Taylor expansion, it follows that

$$\begin{aligned}
K_{1jl}^{\hat{w}} - K_{1jl}^{w} &= \sum_{s=1}^{M} \frac{1}{s!} K_{1jl}^o k_1^{(s)}\left(\frac{V_j - V_l}{h_1}\right)\left[\frac{\hat{V}_j - \hat{V}_i - (V_j - V_i)}{h_1}\right]^s \\
&\quad + \frac{1}{(M+1)!} K_{1jl}^o k_1^{(M+1)}(\phi_{jl})\left[\frac{\hat{V}_j - \hat{V}_i - (V_j - V_i)}{h_1}\right]^{(M+1)},
\end{aligned}$$

where $k_1^{(s)}$ is the sth derivative of $k_1(.)$ and ϕ_{jl} is between $(\hat{V}_j - \hat{V}_i)/h_1$ and $(V_j - V_i)/h_1$. Substitute the last equation into the expression for I_2, we have that

$$\begin{aligned}
I_2 &= \frac{1}{n(n-1)^3 h_1^{2d_1} h_2^{d_2}} \sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq j} (Y_i - Y_k)(Y_j - Y_l) K_{1ik}^{w}\left[K_{1jl}^{\hat{w}} - K_{1jl}^{w}\right] K_{2ij} \\
&= \sum_{s=1}^{M} \frac{1}{s!}\Bigg\{\frac{1}{n(n-1)^3 h_1^{2d_1} h_2^{d_2}} \sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq j} (Y_i - Y_k)(Y_j - Y_l) K_{1ik}^{w} K_{1jl}^o k_1^{(s)}\left(\frac{V_j - V_l}{h_1}\right) \\
&\times \left[\frac{\hat{V}_j - \hat{V}_i - (V_j - V_i)}{h_1}\right]^s K_{2ij}\Bigg\} \\
&+ \frac{1}{(M+1)! n(n-1)^3 h_1^{2d_1} h_2^{d_2}} \sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq j} (Y_i - Y_k)(Y_j - Y_l) K_{1ik}^{w} K_{1jl}^o k_1^{(M+1)}(\phi_{jl}) \\
&\times \left[\frac{\hat{V}_j - \hat{V}_l - (V_j - V_l)}{h_1}\right]^{(M+1)} K_{2ij} \\
&\stackrel{def}{=} \sum_{s=1}^{M} \frac{1}{s!} I_2^s + \frac{1}{r+1} I_2^{M+1}.
\end{aligned}$$

First we prove that $I_2^1 = o_p(n^{-1}h_2^{-d_2/2})$. The expression I_2^1 could be written as

$$I_2^1 = \frac{1}{n(n-1)^3 h_1^{2d_1} h_2^{d_2}} \sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq j} (Y_i - Y_k)(Y_j - Y_l) \frac{\hat{V}_j - \hat{V}_l - (V_j - V_l)}{h_1} K_{1ik}^w K_{1jl}^o k_{1ij}^{(1)} K_{2ij}.$$

Its expectation follows $o_p(n^{-1}h_2^{-d_2/2})$ because

$$\begin{aligned}
E[I_2^1] =& \frac{1}{h_1^{2d_1} h_2^{d_2}} E\left\{ [(Y_1 - Y_3)K_{113}^w] \left[(Y_2 - Y_4) \frac{\hat{V}_2 - \hat{V}_4 - (V_2 - V_4)}{h_1} K_{124}^o k_{124}^{(1)} \right] K_{212} \right\} \\
\stackrel{def}{=}& \frac{1}{h_1^{2d_1} h_2^{d_2}} E\Big\{ E_1 \left[(G(W_1) - G(W_3)) K_{113}^w \right] \\
& \times E_2 \left[(G(W_2) - G(W_4)) \frac{\hat{V}_2 - \hat{V}_4 - (V_2 - V_4)}{h_1} K_{124}^o k_{124}^{(1)} \right] K_{212} \Big\} \\
=& \frac{1}{h_1^{2d_1} h_2^{d_2}} O(h_1^{r+d_1}) O(n^{-1/2} h_0^{-d_0/2} h_1^{d_1}) O(h_2^{d_2}) \\
=& O(n^{-1/2} h_0^{-d_0/2} h_1^{r})
\end{aligned}$$

Likewise, we can get that $Var[I_2^1] = o(n^{-1}h_0^{-d_0}h_1^{2r})$. Therefore, $I_2^1 = O_p(n^{-1/2}h_0^{-d_0/2}h_1^r)$. Under Assumption 3.2.4, we have that $I_2^1 = o_p(n^{-1}h_2^{-d_2/2})$. The proofs for s=2...M follows similarly. For the last term I_2^{M+1}, we can derive in similar way that

$$\begin{aligned}
E[I_2^{M+1}] &= \frac{1}{h_1^{2d_1} h_2^{d_2}} O(h_1^{d_1}) O(n^{-(M+1)/2} h_0^{-d_0(M+1)/2} h_1^{-(M+1)}) O(h_2^{d_2}) \\
&= O(n^{-(M+1)/2} h_0^{-d_0(M+1)/2} h_1^{-(M+1+d_1)})
\end{aligned}$$

Similarly, we get that $I_2^{M+1} = o_p(n^{-1}h_2^{-d_2/2})$ under Assumption 3.2.4. And I_3 could be shown to be $o_p(n^{-1}h_2^{-d_2/2})$ is the same way. Therefore, we get that

Lemma 9. *Under Assumption 3.2.1-3.2.4 and H_0^a,*

$$\frac{n h_2^{d_2/2} \tilde{S}_1}{\sqrt{2}\sigma_\Gamma} \to N(0,1).$$

The rest is to show that $\tilde{\sigma}_\Gamma^2$ is a consistent estimator of σ_Γ^2, which is straightforward as $\hat{M}(\hat{W})$ and $\hat{f}(\hat{W})$ are consistent estimators of $M(W)$ and $f(W)$. $\square$

Proof of Theorem 3.2.3

Proof. Let $\hat{\xi}_i = Y_i - W_i\hat{\theta}$. First we study under the null hypothesis the asymptotic distribution of

$$\begin{aligned}
\hat{S}_3 &= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\hat{\xi}_i\hat{\xi}_j K_{ij} \\
&= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[W_i(\theta-\hat{\theta})+\xi_i\right]\left[W_j(\theta-\hat{\theta})+\xi_j\right]K_{ij} \\
&= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[W_i(\theta-\hat{\theta})W_j(\theta-\hat{\theta})K_{ij} + 2\sum_i\sum_{j\neq i}W_i(\theta-\hat{\theta})\xi_j K_{ij} + \sum_i\sum_{j\neq i}\xi_i\xi_j K_{ij}\right] \\
&\stackrel{let}{=} I + II + III
\end{aligned}$$

We want to show that both I and II are $o_p\left(\left(nh^{d/2}\right)^{-1}\right)$ and that III is asymptotically normal. First, look at the expectation of expression I and I^2. We find that

$$E[I] = h^{-d}E\left[W_1(\theta-\hat{\theta})W_2(\theta-\hat{\theta})K_{12}\right] = h^{-d}O_p(n^{-1/2}n^{-1/2}h^dh^2) = o_p\left(\left(nh^{d/2}\right)^{-1}\right),$$

and

$$\begin{aligned}
I^2 &= \frac{1}{n^2(n-1)^2h^{2d}}\sum_i\sum_{j\neq i}\sum_{i'}\sum_{j'\neq i'}W_i(\theta-\hat{\theta})_j(\theta-\hat{\theta})K_{ij}W_{i'}(\theta-\hat{\theta})W_{j'}(\theta-\hat{\theta})K_{i'j'} \\
&\stackrel{let}{=} A_1 + A_2 + A_3 + A_4 + A_5,
\end{aligned}$$

where A_1 is the sum when none of these four indexes are equal to each other, A_2 is when $i = i'$, $j \neq j'$, A_3 is when $i = j'$, $j \neq i'$, A_4 is when $i = i'$, $j = j'$ and A_5 is when $i = j'$, $j = i'$. It is easy to calculate $E[I^2] = o_p\left(\left(nh^{d/2}\right)^{-2}\right)$ since

$$\begin{aligned}
E[A_1] &= \frac{1}{h^{2d}}E\left[W_1(\theta-\hat{\theta})W_2(\theta-\hat{\theta})K_{12}\right]E\left[W_3(\theta-\hat{\theta})W_4(\theta-\hat{\theta})K_{34}\right] = o_p\left(\left(nh^{d/2}\right)^{-2}\right), \\
E[A_2] = E[A_3] &= \frac{1}{nh^{2d}}E\left[W_1(\theta-\hat{\theta})W_2(\theta-\hat{\theta})K_{12}W_1(\theta-\hat{\theta})W_3(\theta-\hat{\theta})K_{13}\right] = o_p\left(\left(nh^{d/2}\right)^{-2}\right), \\
E[A_4] = E[A_5] &= \frac{1}{n(n-1)h^{2d}}E\left[W_1^2(\theta-\hat{\theta})W_2^2(\theta-\hat{\theta})K_{12}^2\right] = o_p\left(\left(nh^{d/2}\right)^{-2}\right).
\end{aligned}$$

Therefore, $I = o_p\left(\left(nh^{d/2}\right)^{-1}\right)$. Similarly, we could show that $II = o_p\left(\left(nh^{d/2}\right)^{-1}\right)$. Next, we study the asymptotic distribution of III using properties of the U-statistic stated in Lemma B.4 in Fan and Li (1996) which is a generalization of Hall (1984).

$$U_n = h^d II = \frac{2}{n(n-1)} \sum_i \sum_{j<i} P_n(\zeta_i, \zeta_j)$$

where $p(\zeta_i, \zeta_j) = \xi_i \xi_j K_{ij}$. It is easy to see that $E[p(\zeta_i, \zeta_j)] = 0$, $E[p(\zeta_i, \zeta_j)|\zeta_j] = 0$ and the limiting condition in Hall (1984) is satisfied.

$$\begin{aligned}
E\left[P_n^2(\zeta_i, \zeta_j)\right] &= E\left[\xi_1^2 \xi_2^2 K_{12}^2\right] = E\left[\sigma^2(X_1)\sigma^2(X_2)K_{12}^2\right] \\
&= \int f(x_1)f(x_2)\sigma^2(x_1)\sigma^2(x_2)K_{12}^2 dx_1 dx_2 = h^d \int f^2(x_1)\sigma^4(x_1)dx_1 \int K^2(s)ds \\
&= h^d CE[f(X_1)\sigma^4(X_1)] = h^d CE[\xi_1^2 E[\xi_1^2|X]f(X)] = h^d \sigma_\Gamma^2.
\end{aligned}$$

Then we get that $nh^{d/2}II = \frac{1}{h^{d/2}} nU_n \rightarrow N(0, 2\sigma^2)$. Therefore,

$$\frac{nh^{d/2}\hat{S}_3}{\sqrt{2}\sigma_\Gamma} \rightarrow N(0,1)$$

Then the rest is to show that $\tilde{S}_3 = \hat{S}_3 + o_p\left(\left(nh^{d/2}\right)^{-1}\right)$ and that $\tilde{\sigma}^2 - \sigma^2 = o_p(1)$. The former follows straightforwardly from the fact that $\hat{\rho} - \rho = O_p(n^{-1/2})$ and the latter could be derived using the same argument as in the last part of the proof for Theorem 3.2.1. $\square$

Bibliography

[1] Jason Abrevaya. The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics*, 26:247–257, 2001.

[2] Jason Abrevaya and Shu Shen. Estimation of censored panel-data models with slope heterogeneity. 2010.

[3] Hyungtaik Ahn. Nonparametric two-stage estimation of conditional choice probabilities in a binary choice model under uncertainty. *Journal of Econometrics*, 67:337–378, 1995.

[4] Sule Alan, Bo E. Honoré, and Soren Leth-Petersen. Estimation of panel data models with two-sided censoring. SSRN working paper 1325276, 2008.

[5] Rob Alessie, Stefan Hochguertel, and Arthur van Soest. *Household portfolios*, chapter Household portfolios in the Netherlands, pages 341–388. MIT Press, 2002.

[6] Gordon Anderson. Nonparametric tests of stochastic dominance in income distributions. *Econometrica*, 64:1183–1193, 1996.

[7] Donald W. K. Andrews and Xiaoxia Shi. Inference based on conditional moment inequalities. 2010.

[8] Joshua Angrist, Victor Chernozhukov, and Ivan Fernandez-Val. Quantile region under misspecification, with an appication to the u.s. wage structure. *Econometrica*, 74:539563, 2006.

[9] Joshua D. Angrist and William N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88:450–477, 1998.

[10] Yacine. At-Sahalia, Peter J. Bickel, and Thomas M Stoker. Goodness-of-fit tests for regression using kernel methods. *Journal of Econometrics*, 105:363–412, 2001.

[11] Garry F. Barrett and Stephen G. Donald. Consistent tests for stochastic dominance. *Econometrica*, 71:71–104, 2003.

[12] Vijia S. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2:95–121, 1975.

[13] Patrick Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 1999.

[14] Richard W. Blundell and James L. Powell. *Advances in Economics and Econometrics Theory and Applications*, chapter Endogeneity in Nonparametric and Semiparametric Regression Models. Cambridge University Press, 2003.

[15] Richard W. Blundell and James L. Powell. Endogeneity in semiparametric binary response models. *Review of Economic Studies*, pages 655–679, 2004.

[16] Gary Chamberlain. *Handbook of econometrics, vol. 2*, chapter Panel data, pages 1247–1318. North Holland, Amsterdam, 1984.

[17] Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75:1243–1284, 2007.

[18] Christopher Cornwell, Peter Schmidt, and Robin C. Sickles. Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46:185–200, 1990.

[19] M. Csorgo and P. Revesz. *Strong Approximation in Probability and Statistics.* Academic Press, 1981.

[20] Russel Davidson and Jean-Yves Duclos. Statistical inference for stochastic dominance and for the measurment of poverty and inequliaty. *Econometrica*, 68:1435–1464, 2000.

[21] Miguel A. Delgado and Juan Carlos Escanciano. Distribution-free tests of stochastic monotonicity. 2010.

[22] Stephen Donald and Yuchin Hsu. Testing for stochastic dominance in treatment effects. 2009.

[23] Stephen Donald and Yuchin Hsu. Improving the power of tests of stochastic dominance. 2010.

[24] Stephen G. Donald, Guido W. Imbens, and Whitney K. Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117:55–93, 2003.

[25] R. M. Dudley. Weak convergence of probabilities on nonseperable metric spaces and empirical measures on euclidean spaces. *Illinois Journal of Mathametics*, 10:109–126, 1966.

[26] R. M. Dudley. Central limit theorems for empirical measures. *The Annals of Probability*, 6:899–929, 1978.

[27] Yanqin Fan and Qi Li. Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64(4):865–90, July 1996.

[28] Sergio Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75:259–276, 2007.

[29] Pedro L. Gozalo. A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, 9:451–477, 1993.

[30] Peter Hall and Joel L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*, 33:2904–2929, 2005.

[31] Peter Hall and Adonis Yatchew. Unified approach to testing functional hypotheses in semiparametric contexts. *Journal of Econometrics*, 127:225–252, 2005.

[32] Bruce E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64:413–430, 1996.

[33] Wolfgang Hardle and Thomas M. Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84:986–995, 1989.

[34] Wolfgang Karl Hardle and Enno Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21:1926–1947, 1993.

[35] Jerry A. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–1271, 1978.

[36] James Heckman and V. Joseph Hotz. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *Journal of the American Statistical Association*, 84:862–874, 1989.

[37] Stephen Hochguertel. Precautionary motives and portfolio decisions. *Journal of Applied Econometrics*, 18:61–77, 2003.

[38] Bo E. Honoré. Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60:533–565, 1992.

[39] Joel Horowitz and Wolfgang Hardle. Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association*, 91:1632–1640, 1996.

[40] Joel L. Horowitz. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74:521–538, 2006.

[41] Lajos Horvath and Brian Yandell. Asymptotics of condtional empirical processes. *Journal of Multivariate Analysis*, 26:184–206, 1988.

[42] Lajos Horvatha, Piotr Kokoszkab, and Ricardas Zitikisc. Testing for stochastic dominance using the weighted mcfadden-type statistic. *Journal of Econometrics*, 133:191205, 2006.

[43] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58:71–120, 1993.

[44] Guido W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29, 2004.

[45] Guido W. Imbens and Whitney K. Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77:1481–1512, 2009.

[46] Yuichi Kitamura, Gautam Tripathi, and Hyungtaik Ahn. Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72:16671714, 2004.

[47] Roger W. Klein and Richard H. Spady. An efficient semiparametric estimator for discrete choice models. *Econometrica*, 61:387–422., 1993.

[48] Roger Koenker and Beum J. Park. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71:265–283, 1996.

[49] Roger Koenker and Zhijie Xiao. Inference on the quantile regression process. *Econometrica*, 70:1583–1612, 2002.

[50] Sokbae Lee, Oliver Linton, and Yoon-Jae Whang. Testing for stochastic monotonicity. *Econometrica*, 77:585–602, 2009.

[51] Sokbae Lee and Yoon-Jae Whang. Noparametric tests of condtional treatment effects. 2009 2009.

[52] Oliver Linton, Esfandiar Maasoumi, and Yoon-Jae Whang. Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, 72:735–765, 2005.

[53] Oliver Linton, Kyungchul Song, and Yoon-Jae Whang. An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154:186–202, 2010.

[54] lizbar A. Nadaraya. On estimating regression. *Theory of Probability and Its Applications*, 9:141–142, 1964.

[55] Shelly Lundberg and Elaina Rose. The effects of sons and daughters on men's labor supply and wages. *The Review of Economics and Statistics*, 84:251–268, 2002.

[56] Brian McCaig and Adonis Yatchwe. International welfare comparisons and noparametric testing of multivariate stochastic dominance. *Journal of Applied Econometrics*, 22:951–969, 2007.

[57] Daniel McFadden. *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, chapter Testing for Stochastic Dominance. Springer, 1989.

[58] Robert C. Merton. Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics*, 51:247–257, 1969.

[59] Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413, 1971.

[60] Yair Mundlak. On the pooling of time series and cross section data. *Econometrica*, 46:69–85, 1978.

[61] Georg Neuhaus. On weak convergence of stochastic processes with multidimensional time parameter. *The Annals of Mathematical Statistics*, 42:1285–1295, 1971.

[62] Whitney Newey, James Powell, and Francis Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67:565–603, 1999.

[63] Whitney K. Newey. Kernel estimation of partial means and a general variance estimation. *Econometric Theory*, pages 233–253, 1994.

[64] Whitney K. Newey and James L. Powell. Instrumental variables estimation of nonparametric models. *Econometrica*, 71:1565–1578, 2003.

[65] Adrian Pagan and Aman Ullah. *Nonparametric Econometrics*. Cambridge University Press, 1999.

[66] Leslie E. Papke and Jeffrey M. Wooldridge. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145:121–133, 2008.

[67] Solomon W. Polachek and Moon-Kak Kim. Panel estimates of the gender earnings gap: individual-specific intercept and individual-specific slope models. *Journal of Econometrics*, 61:23–42, 1994.

[68] James L. Powell. Least absolute deviations estimation for the censored regressionmodel. *Journal of Econometrics*, 25:303–325, 1984.

[69] James L. Powell. *Handbook of Econometrics*, chapter Estimation of Semiparametric Models. Amsterdam: ElsevierNorth-Holland, 1994.

[70] James L. Powell, James H. Stock, and Thomas M. Stoker. Semiparametric estimation of index coefficients. *Econometrica*, 57:1403–1430, 1989.

[71] James L. Powell and Thomas M. Stoker. Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics*, 75:291–316, 1996.

[72] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in Fortran 90*. Cambridge University Press, 1996.

[73] Xiaohong Chen Richard Blundell and Dennis Kristensen. Semi-nonparametric iv estimation of shapeinvariant engel curves. *Econometrica*, 75:1613–1669, 2007.

[74] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

[75] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:642–669, 1956.

[76] Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51:239–246, 1969.

[77] Bernard Walter Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.

[78] C. J. Stone. *Recent Advances in Statistics*, chapter Optimal Univorm Rate of Convergence for Nonparametric Estimators of a Density Function or its Derivatives, pages 393–406. Academic Press, New York, 1983.

[79] Winfried Stute. Conditional empirical processes. *The Annals of Statistics*, 14:638–647, 1986.

[80] Winfried Stute. On almost sure convergence of conditional empirical distribution functions. *The Annals of Probability*, 14:891–901, 1986.

[81] Winfried Stute and Lixing Zhu. Nonparametric checks for single index models. *Annals of Statistics*, 33:1048–1083, 2005.

[82] Aad W. Van Der Vaart and Jon. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistic*. New York: Springer-Verlag, 1996.

[83] H. D. Vinod and Aman Ullah. Flexible production estimation by noparametric kernel estimators. In *Advanced in Econometrics: Nonparametric and Robust Statistical Inference, Vol 7,*. London, JAI Press, 1988.

[84] Geoffrey S. Watson. Smooth regression analysis. *Sankhya, Series A*, 26:359–372, 1964.

[85] Yoon-Jae Whang and Donald W. K. Andrews. Tests of specification for parametric and semiparametric models. *Journal of Econometrics*, 57:277–318, 1993.

[86] Jeffrey M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT Press, 2002.

[87] Jeffrey M. Wooldridge. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *The Review of Economics and Statistics*, 87:385–390, 2005.

[88] Jeffrey Marc Wooldridge. Control function and related methods. Whats New in Econometrics, NBER Lecture Notes, 2007, 2007.

[89] Adonis Yatchew. Nonparametric regression tests based on an infinite dimensional least squares procedure. *Econometric Theory*, 8:435–451, 1992.